THE GENETICS OF NONTRADITIONAL GLYCEMIC BIOMARKERS OF TYPE 2 DIABETES

by

Stephanie J. Loomis, MPH

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

May, 2018

Abstract

Type 2 diabetes is a major public health problem that affects over 10% of the US adult population. It is associated with substantially increased risks of mortality and serious clinical outcomes such as heart disease, stroke, kidney disease and retinopathy. Diabetes is defined by hyperglycemia, or elevated glucose concentrations in the blood, which are commonly measured by fasting glucose and hemoglobin A1c (HbA1c), but these have limitations. As a result, nontraditional glycemic biomarkers, fructosamine, glycated albumin and 1,5-anhydroglucitol (1,5-AG) are gaining interest. While it is established that genetics play a role in type 2 diabetes, fasting glucose, and HbA1c, the genetics of fructosamine, glycated albumin, and 1,5-AG have not been well explored. This dissertation sought to determine the amount of variation in each biomarker due to genetics through heritability estimation, and to determine the specific genetic variants associated with each biomarker though genome wide association study (GWAS) analysis, multivariate phenotype analysis, and exome sequencing analysis.

Heritability estimates showed a substantial portion of fructosamine, glycated albumin and 1,5-AG variation was due to genetics, which is likely comprised of both common and rare variants. GWAS identified common variants associated with fructosamine and glycated albumin including a known diabetes variant and a likely nonglycemic variant. Exome sequencing did not identify variants associated with fructosamine and glycated albumin, but multivariate phenotype analysis identified a potentially interesting region in a gene that alters bilirubin levels that may affect fructosamine in a nonglycemic manner. Exome sequencing identified rare, coding variants with large effect size in a glucose transporting gene associated with 1,5-AG

which inform the biology and may impact the clinical interpretation of 1,5-AG. Analyzing the genetics of nontraditional glycemic biomarkers of type 2 diabetes has increased the understanding of these biomarkers, including their underlying biology, and may aid in decisions about their clinical implementation.

Dissertation Readers

Priya Duggal, PhD, MPH (Advisor)
Associate Professor of Epidemiology and International Health

Elizabeth Selvin, PhD, MPH
Professor of Epidemiology and Medicine

Rasika Mathias, ScD
Associate Professor of Medicine

Ni Zhao, PhD
Assistant Professor of Biostatistics

Acknowledgements

I would like to thank a number of people who have helped me along the way to pursuing a PhD. Firstly, I would like to thank Priya Duggal, who has been an excellent advisor. She has given me so much of her time and helped guide me through each step of the PhD process. She has challenged me to move beyond the methods and results to really grasp the "but why does it matter?" aspect of my work. I would also like to thank Liz Selvin, for venturing into the genetics world as my co-advisor. She gives so much to her students and is incredibly patient with multiple rounds of revisions to make sure my work is the best it can be.

I would also like to thank the other members of my dissertation committee, Joe Coresh and Adrienne Tin for their helpful insight. My coauthors on this work have given thoughtful and useful comments. In particular, Anna Köttgen has provided guidance on much of this work. Even from across the Atlantic, she is ready with prompt replies with knowledgeable answers to may many emails. Debashree Ray has been especially helpful with the multivariate phenotype analysis project. I would also like to thank my dissertation readers, Priya Duggal, Liz Selvin, Rasika Mathias and Ni Zhou.

The members of my epidemiology PhD doctoral cohort have been such a wonderful aspect of my time at Hopkins. They are great friends and colleagues, and have been so supportive, listening to practice talks, preparing for oral exams, and letting me randomly walk into their offices to bounce ideas off of them. I would like to thank the faculty and students in the genetic epidemiology track, who are incredibly knowledgeable and were always willing to do discuss methods and share code. The members of Priya's

research group have been so helpful in providing excellent feedback, particularly at the early stages of this work.

Finally, I would like to thank my incredibly supportive family and friends, who have been there for the good, bad, and everything in-between during this PhD process.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Type 2 diabetes is an important public health concern

Type 2 diabetes is a major public health problem facing the US and countries around the world. The current prevalence of diabetes is 10% in the US,[1] , similar to the age-standardized worldwide prevalence (9% in men, 7.9% in women) and reaches over 25% in Polynesia and Micronesia, with a high in American Samoa of more than 30%.[2] 300 million individuals have been diagnosed with diabetes worldwide, and this number is expected to increase to over 550 million by 2030.[3]

Diabetes occurs when glucose levels in the blood are elevated (hyperglycemia), which is caused by a combination of beta cell failure and insulin resistance.[4] In a healthy individual, excess glucose in the blood (e.g., as a result of a recent meal) triggers pancreatic beta cells to excrete insulin, a hormone which facilitates glucose uptake from the blood into tissues such as muscle, liver and fat, where it is either used for energy or stored for future use. In an individual with type 2 diabetes, however, pancreatic beta cells fail to adequately excrete insulin and/or the glucose-absorbing cells fail to uptake glucose in response to insulin (insulin resistance) which results in chronic hyperglycemia.[5] Diabetes leads to major health complications including blindness, loss of limb, kidney disease, heart disease, stroke and death.[4,6,7]

1.2 Environmental and genetic factors increase risk of type 2 diabetes

Many behavioral and environmental factors affect the risk of type 2 diabetes, including obesity, lack of physical activity, age, and race.[4] Genetics also plays a role in

type 2 diabetes risk. Heritability, or the proportion of variance in diabetes due to genetics, has been estimated from 20% to 80%, with the wide range likely due to differences in diabetes definitions, populations, heritability estimation methods, sample types (twins vs other family members), and sample size. A recent study combining 7 twin cohorts including 34,166 twin pairs estimated heritability of type 2 diabetes to be 72%.[8] This suggests a substantial genetic role in diabetes risk.[8-11] Genome-wide association studies (GWAS) and sequencing studies have identified over 100 genetic variants associated with diabetes, using diverse and large samples (up to 150,000 individuals).[12,13] The majority of the variants identified thus far are common (minor allele frequency≥5%) and have small effect sizes (**Figure 1.1**). All of these variants, however, only account for approximately 10% of the heritability of diabetes, indicating that much work is still to be done to understand diabetes genetics.[12]

1.3 Type 2 diabetes results in elevated blood glucose levels, measured by biomarkers

The hallmark of type 2 diabetes, hyperglycemia, can be characterized by different biomarkers. The most commonly used clinical biomarkers of hyperglycemia are fasting glucose and hemoglobin A1c (HbA1c), both recommended for screening and diagnosis by the American Diabetes Association, along with less commonly used oral glucose tolerance test.[4] Fasting glucose is a measure of glucose concentration in the blood following an 8 hour fast and represents glucose levels at a fixed time point. HbA1c is a test that measures the percent of hemoglobin in the blood which is glycated. HbA1c is formed by a nonenzymatic binding of glucose to hemoglobin contained within red blood cells. It represents average blood glucose levels over the previous 2-3 months.[14]

Fasting glucose and HbA1c both have limitations in their abilities to accurately reflect diabetes status. Fasting glucose is burdensome on the patient (requiring an 8 hour fast), has high pre-analytic variability, moderate intra-individual variability, and can be affected by factors such as recent illness, exercise or stress. HbA1c levels may be affected by erythrocyte or hemoglobin related factors such as anemia and rare hemoglobin variants.[15-17]

1.4 Nontraditional biomarkers of hyperglycemia may overcome some limitations of traditional measures and are of growing clinical interest

In light of certain limitations of fasting glucose and HbA1c, additional biomarkers of hyperglycemia have been proposed for use in diabetes care: fructosamine, glycated albumin and 1,5-anhydroglucitol (1,5-AG).[14,17] Fructosamine and glycated albumin are both ketoamines, formed by nonenzymatic binding of serum protein to glucose.[18] Fructosamine is a measure of the concentration of total serum protein bound to glucose, while glycated albumin is a measure of the percent of albumin (the most prevalent serum protein) that is glycated. Both represent average blood glucose over the previous 2-3 weeks.[18] 1,5-AG is a carbohydrate that is structurally similar to glucose and is consumed though certain foods including soybeans, rice, bread and beef.[19-21] In normoglycemic conditions, 1,5-AG is filtered through the kidney, resulting in stable concentrations in the blood. In hyperglycemic conditions above the renal threshold for absorption of glucose ($\geq$180 mg/dL), glucose outcompetes 1,5-AG for reabsorption in the kidney, leading to lower levels of 1,5-AG in the blood.[19-22] 1,5-AG represents spikes in blood glucose levels occurring in the previous 1-2 weeks.[19-21] All three biomarkers are associated with

diabetes and diabetes-related outcomes similarly to fasting glucose and HbA1c, and can predict diabetes after controlling for fasting glucose or HbA1c.[23-32]

While these nontraditional biomarkers are not commonly used clinically in the US,[15] fructosamine is recommended as an alternative to HbA1c in individuals with known erythrocyte disorders by organizations in India, Australia and the UK.[17] Glycated albumin is regularly used clinically to monitor short-term changes in glycemic control for diabetes in China, Japan and South Korea (age-standardized diabetes prevalence in women and men >18 years: 7.6%, 9.9%; 5%, 8.4%; 6.7%, 9.3%, respectively)[2] and was recently cleared by the FDA for clinical use in the US.[17,33]

Similar to HbA1c, which is an indirect measure of hyperglycemia, fructosamine and glycated albumin also have potential limitations based on the non-glucose dependent portion of the molecule. Factors that affect albumin metabolism and serum protein levels such as kidney, liver and thyroid disease are known to alter fructosamine and glycated albumin levels.[34,35] 1,5-AG can be affected by diet and kidney impairment.[20]

1.5 Genetics of traditional hyperglycemia biomarkers have been well-studied

The genetics of fasting glucose and HbA1c have been well studied. Multiple GWAS including one with over 133,000 participants from the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) found 36 variants associated with fasting glucose.[36] Additionally, a recent paper incorporating approximately 160,000 samples from African, European, and Asian ancestries increased the number of HbA1c loci to 60.[37]

Interestingly, despite fasting glucose and HbA1c both being measures of hyperglycemia, only a small number of the genetic variants associated with these biomarkers overlap (**Figure 1.2**) and the genetic correlation is not strong (r=0.41).[38] This is likely due to the differences in the biology of the biomarkers and the glucose states they represent. Wheeler et. Al. determined that 22 of 60 HbA1c variants were associated with erythrocytic factors not related to diabetes status but to the erythrocytes in which HbA1c resides, while 19 variants were glucose-related "glycemic" variants (**Figure 1.2**).[37] In addition, each biomarker represents different aspects of glycemic states which may also account for the observed differences: HbA1c reflects average blood glucose across 2-3 months, while fasting glucose is a measure of instantaneous hepatic glucose output.

Several studies of HbA1c genetics have shown that nonglycemic genetic variants can alter HbA1c levels enough to reclassify diabetes status.[37,39,40] The lowered HbA1c levels due to nonglycemic genetic factors has the potential to underdiagnose a substantial number of people (650,000 African American adults in the US alone by one estimate)[37] screened for diabetes by HbA1c, highlighting the importance of investigating glycemic biomarker genetics.

1.6 Little is known of the genetics of nontraditional hyperglycemia biomarkers

While much work has been done to investigate HbA1c and fasting glucose genetics, little is known about the genetics of nontraditional hyperglycemia biomarkers. Heritability, has been estimated for 1,5-AG measured using an untargeted assay. This study estimated approximately 60% of variation in 1,5-AG was due to genetics.[41,42]

Heritability has not been estimated for 1,5-AG using a targeted assay, however. We recently identified common genetic variants associated with 1,5-AG through a GWAS in the Atherosclerosis Risk in Communities (ARIC) Study. Seven loci, five of which were novel, in genes involved in glucose transport and carbohydrate metabolism were significantly associated with 1,5-AG,[43] identifying a potentially important pathway involved in diabetes pathophysiology that has not been discovered through genetic studies of diabetes or other glycemic biomarkers (**Figure 1.3, 1.4**). While this GWAS captured common variants associated with 1,5-AG, it was not designed to capture rare variants. Because the genetic architecture of these biomarkers is likely comprised of both common and rare variants, an important portion of the genetic variants associated with 1,5-AG has not been investigated.

Even less is known about the genetics of fructosamine and glycated albumin. No heritability studies, nor studies to identify common or rare variants associated with fructosamine and glycated albumin have been done. Given the known clinical impact of nonglycemic genetics on the ability of HbA1c to accurately reflect diabetes status, and the structure of fructosamine and glycated albumin are similarly impacted by both glucose and nonglucose factors, it is imperative to understand the genetics of nontraditional markers of hyperglycemia to inform their clinical interpretation.

1.7 Dissertation specific aims

The purpose of this dissertation was to investigate the genetics of nontraditional glycemic biomarkers – fructosamine, glycated albumin and 1,5-AG – and to compare them to traditional glycemic biomarkers.

This work served two purposes:

1) to inform the biology of the specific biomarkers under study and potential implications for their clinical use; and

2) to inform the general biology of diabetes

All of the five hyperglycemia biomarkers studied herein (fasting glucose, HbA1c, fructosamine, glycated albumin and 1,5-AG) are correlated, but vary in the strength of their correlations.[25,28] Similarly, the genetic variants associated with the biomarkers do not overlap as much as might be expected given that each is meant to reflect glycemic status. This may be because some of the genetic variants are associated with the nonglycemic aspects of the biomarkers (e.g., erythrocytes in HbA1c), or because the different biomarkers reflect different biologic aspects of type 2 diabetes (e.g., fasting glucose as a measure of hepatic glucose output, 1,5-AG as a measure of glycemic excursions, and HbA1c, fructosamine and glycated albumin as measures of average glucose levels over time). Conversely, variants that are associated with all of the glycemia biomarkers are likely to be more specific to type 2 diabetes. We used the similarities and differences in biomarker genetics to inform biomarker and diabetes biology.

This dissertation was undertaken to address the following research questions:

How much of fructosamine, glycated albumin and 1,5-AG is under genetic control, and how does this compare to HbA1c and fasting glucose?

Aim 1: Determine the narrow-sense and SNP-based heritability of biomarkers of type 2 diabetes (fructosamine, glycated albumin, 1,5-AG, fasting glucose, HbA1c) using genome-wide association (GWAS) data from a single large-population based cohort.

> *Hypothesis 1*: There is a genetic component to nontraditional glycemic markers, which can be quantified and will inform the amount of variation due to genetics for each marker. Comparing narrow-sense and SNP-based heritabilities will inform the genetic architecture of these biomarkers.

What specific variants contribute to the genetic component of fructosamine and glycated albumin? Do these variants overlap with other glycemic biomarkers, and are the variants "glycemic" (diabetes related) or "nonglycemic" (limitation of the test)?

Aim 2: Identify common genetic variants, both independent and shared with other glycemic biomarkers, associated with fructosamine and glycated albumin across the genome.

> *Hypothesis 2:* Common variants are associated with fructosamine and glycated albumin, which may be glycemic or nonglycemic.

Aim 3: Identify rare, exonic variants associated with fructosamine and glycated albumin using a univariate approach, and identify additional common and rare variants using multivariate phenotype analyses.

Hypothesis 3: Rare variants are associated with fructosamine and glycated albumin. In addition, the increased power of multivariate phenotype analysis over single phenotype analysis will identify additional variants associated with fructosamine and glycated albumin which were not identified in the single phenotype analysis.

What specific variants contribute to the genetic component of 1,5-AG? Do these variants overlap with variants associated with other glycemic biomarkers, and are the variants "glycemic" (diabetes related) or "nonglycemic" (limitation of the test)?

Aim 4: Identify rare, exonic genetic variants associated with 1,5-AG that may contribute to its overall genetic architecture.

Hypothesis 4: There are rare variants that underlie the genetic architecture of 1,5-AG. These rare variants may reflect glycemic and nonglycemic genetic control of nontraditional glycemic markers.

Discerning the genetics of fructosamine, glycated albumin and 1,5-AG sought to inform these markers' ability to accurately reflect blood glucose levels, with potential implications for their use in the setting of diabetes care and inform diabetes-related mechanisms not captured by other measures.

1.8 Tables and Figures

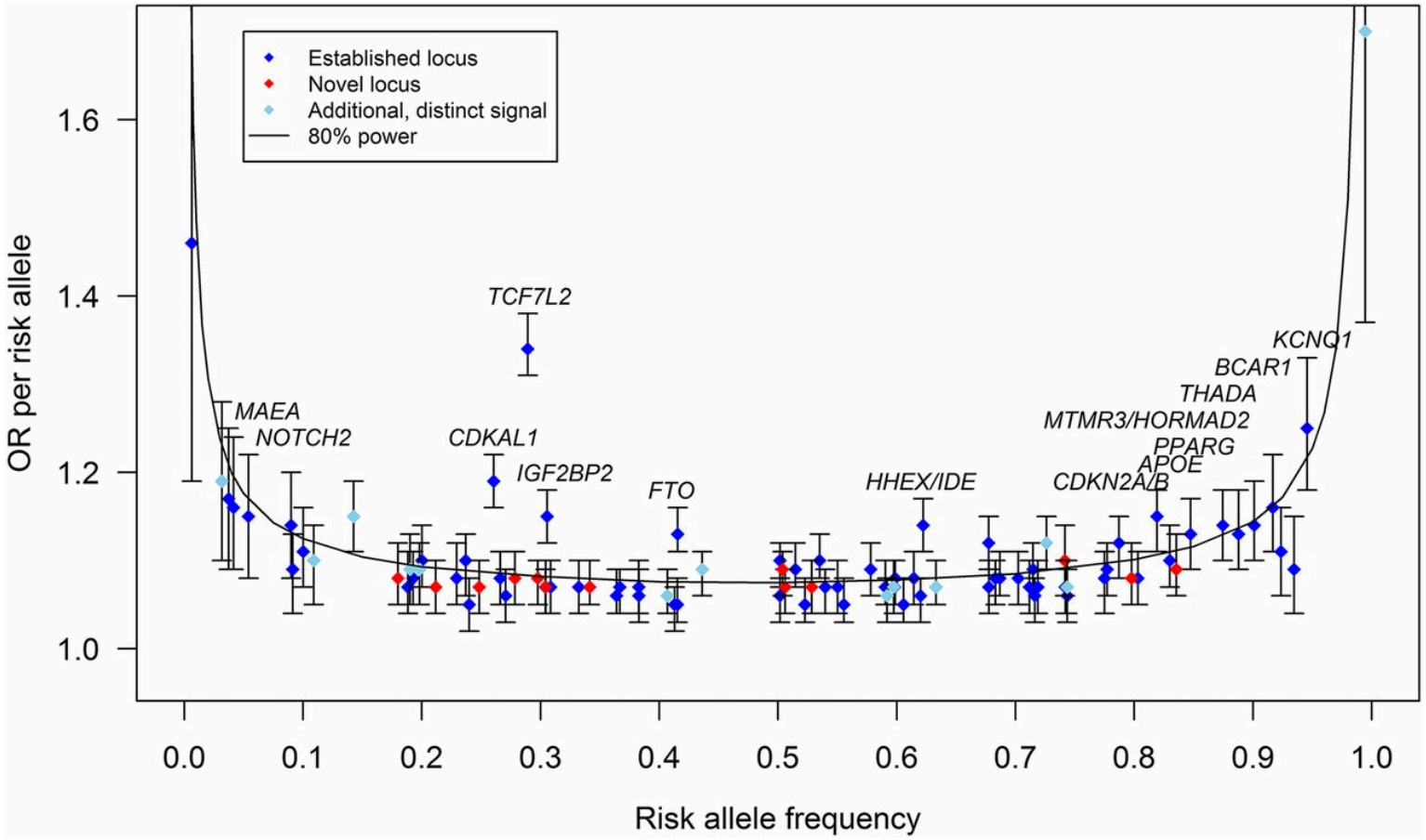


Figure 1.1 Variants discovered thus far for type 2 diabetes by effect size and allele frequency (figure from Scott 2017 Diabetes, used with permission.)[13]

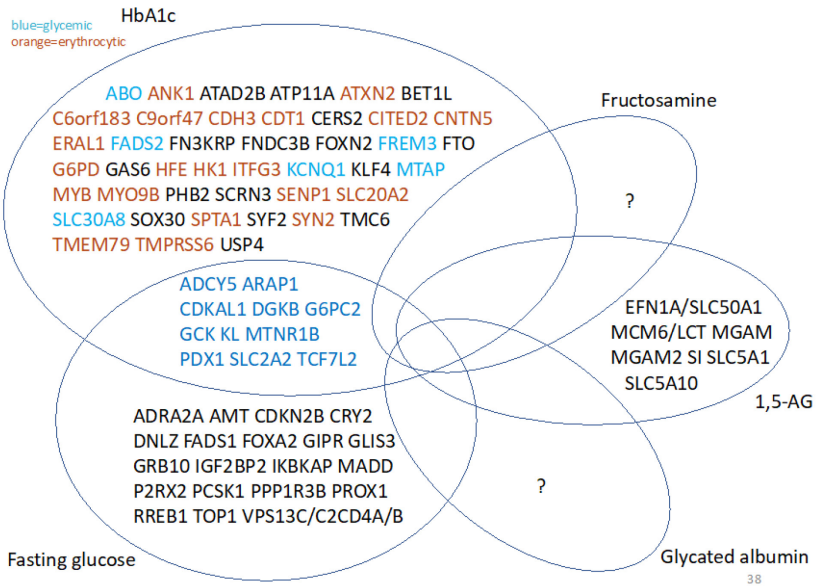


Figure 1.2. Genes with known genetic variants associated with hyperglycemia biomarkers. Blue indicates "glycemic" genes and orange indicates "erythrocytic" genes as determined by Wheeler et al.[37]

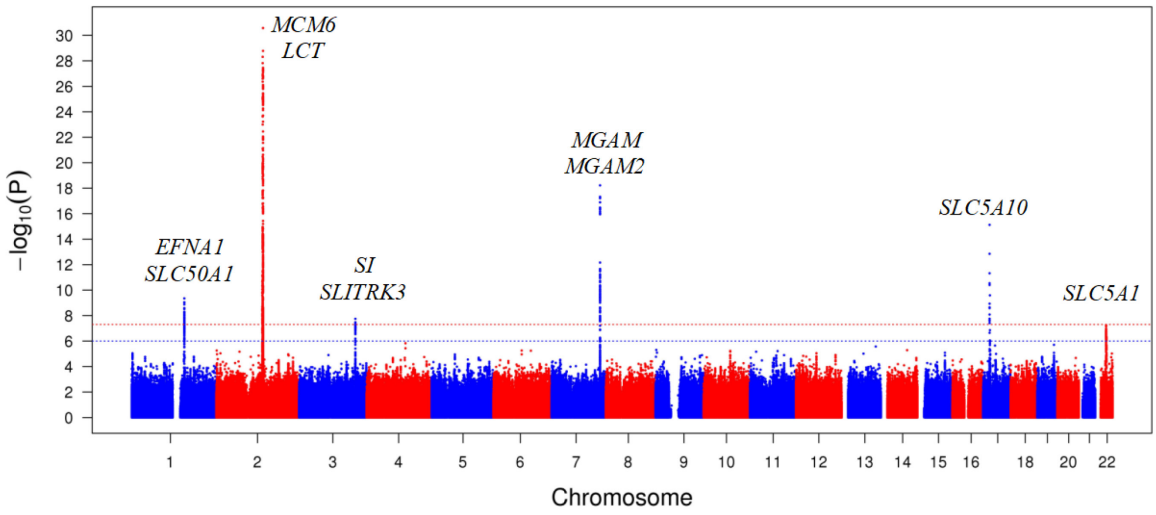


Figure 1.3. Manhattan plot for GWAS of 1,5-anhydroglucitol in European American participants from the ARIC study.[43]

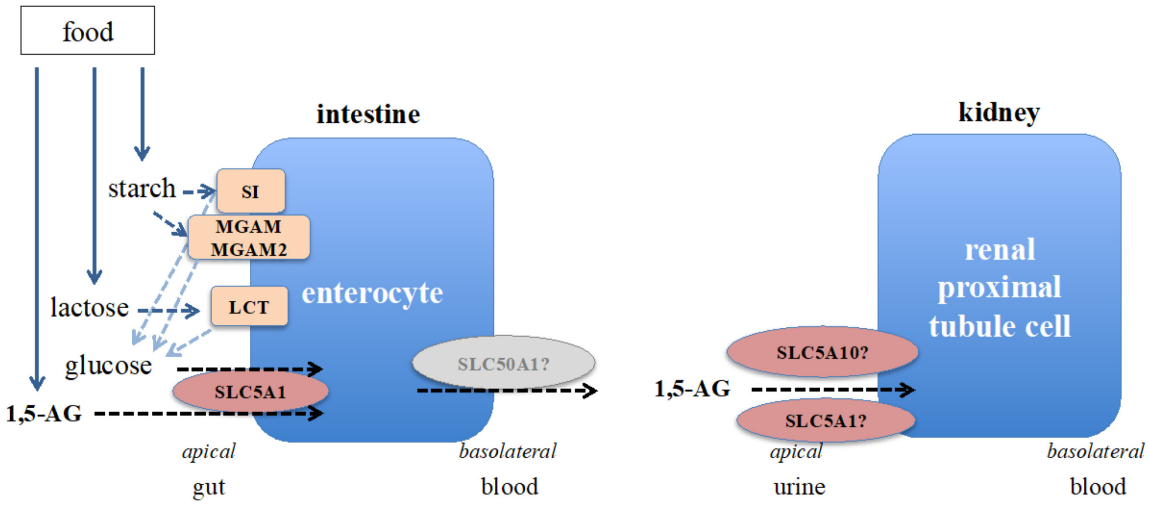


Figure 1.4. Glucose metabolism as a common and biologically plausible theme among the genes mapping into the identified loci. This figure shows their role in intestinal carbohydrate digestion as well as glucose and 1,5-AG reabsorption in gut and kidney.[43]

Chapter 2: Heritability analysis of nontraditional hyperglycemia biomarkers in the Atherosclerosis Risk in Communities (ARIC) Study

Authors:

Stephanie J. Loomis[1], Adrienne Tin[1], Josef Coresh[1,2], Eric Boerwinkle[3], James S. Pankow[4], Xiaoming Liu[3], Anna Köttgen[1,5], Elizabeth Selvin[1,2] *, Priya Duggal[1]*

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

[2] Welch Center for Prevention, Epidemiology, & Clinical Research, The Johns Hopkins University, Baltimore MD

[3]Department of Epidemiology, The University of Texas Health Science Center at Houston School of Public Health at Houston, Houston, TX

[4]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN

[5]Institute of Genetic Epidemiology, Medical Center and Faculty of Medicine - University of Freiburg, Freiburg, Germany

*contributed equally to this work

2.1 Abstract

Introduction: Nontraditional glycemic biomarkers including fructosamine, glycated albumin and 1,5-anhydroglucitol (1,5-AG) are potential alternatives or compliments to traditional measures of hyperglycemia. Genetic variants are associated with these biomarkers, but the heritability, or extent to which genetics control their variation, is not known.

Methods: Narrow-sense, SNP-based and bivariate heritabilities were estimated for traditional glycemic biomarkers (fasting glucose, HbA1c), and nontraditional biomarkers (fructosamine, glycated albumin, 1,5-AG) among white participants in the Atherosclerosis Risk in Communities (ARIC) Study (N=400 first-degree relatives, N=5,575 unrelated individuals).

Results: Narrow-sense heritabilities (representing heritability from the entire genome) for nontraditional biomarkers were substantial (0.44 - 0.55) and comparable to HbA1c (0.34); the fasting glucose estimate was nonsignificant. SNP-based heritabilities (representing heritability from common variants) were lower than narrow-sense heritabilities for all biomarkers. Bivariate heritabilities showed shared genetics between fructosamine and glycated albumin (0.46 narrow-sense, 1.00 SNP-based) and glycated albumin and 1,5-AG (0.50 narrow-sense, 0.47 SNP-based).

Conclusions: Genetic factors contribute to a considerable proportion of the variance of fructosamine, glycated albumin and 1,5-AG and a portion of this heritability likely comes from common variants.

2.2 Introduction

Type 2 diabetes is a major public health problem that affects over 10% of the US adult population and is associated with substantially increased risks of mortality and serious clinical outcomes such as heart disease, stroke, chronic kidney disease and retinopathy.[1,2] Diabetes is defined by hyperglycemia, or elevated glucose concentrations in the blood. Fasting glucose and hemoglobin A1c (HbA1c) are the most common biomarkers used for screening and diagnosis of diabetes, but have limitations. Fasting glucose requires substantial patient preparation (i.e., an eight-hour fast), has high pre-analytic variability, is acutely affected by factors such as recent physical activity or illness, and has moderate intra-individual variability. HbA1c is less affected by these factors, but the interpretation of HbA1c can be problematic in the setting of altered red blood cell turnover or changes in hemoglobin, factors due to characteristics of the biomarker and unrelated to circulating glucose.[2-5] The limitations of traditional measures of hyperglycemia have led to a growing interest in nontraditional biomarkers including fructosamine, glycated albumin and 1,5-anhydroglucitol (1,5-AG).[3,6]

Fructosamine, glycated albumin and 1,5-AG are indirect measures of blood glucose levels. Fructosamine and glycated albumin are both biomarkers where glucose is bound to protein. Fructosamine is glucose bound to serum total protein, and glycated albumin is glucose bound to serum albumin. The majority of serum protein is comprised of albumin, thus there are expected similarities between these two biomarkers and they represent the average blood glucose over the previous ~2-3 weeks.[7]

1,5-AG is a molecule structurally similar to glucose and is consumed through food. During hyperglycemic conditions, when glucose exceeds the renal threshold, glucose is preferentially reabsorbed from urine by the kidney, leading to excretion of 1,5-

AG in the urine and a reduction of serum 1,5-AG levels. Blood 1,5-AG concentrations represent glycemic excursions above the renal threshold over the previous 1-2 weeks.[8-10]

Heritability, the proportion of variance in a phenotype that can be attributed to genetics, is population specific, and is affected by the relative genetic and environmental impacts on the phenotype. Previous studies in various populations have estimated the narrow-sense heritability of fasting glucose to range from 0.30 to 0.70, and HbA1c to range from 0.20 to 0.75.[11-19] Recent studies evaluating hundreds of metabolites using non-targeted assays have estimated the heritability of 1,5-AG to be 0.61 to 0.63[11,20] in population-based studies. To date, no study has estimated the heritability of fructosamine or glycated albumin. Quantifying the genetic contribution of these biomarkers will inform the extent to which genetics may play a role in these non-traditional biomarkers, and determine if they are comparable to traditional diabetes biomarker (fasting glucose and HbA1c) heritabilities.

An underlying assumption of heritability is that if a trait is heritable, individuals who are more closely related will have more similar phenotypes than those who share less genetics (i.e., are distantly related or unrelated). Traditional heritability methods use closely related individuals (first- and second-degree relatives) and infer the degree of shared genetics based on family structure. These methods provide estimates of narrow-sense heritability (h^2), or the proportion of variance in a phenotype passed down from parents to offspring. In newer SNP-based (h^2_{SNP}) heritability methods, the amount of shared genetics among unrelated individuals can be estimated using measured genotypes, because all members of a species have a common ancestor, they share a small amount of their genomes.[21,22] Genotyped variants are often available from genome-wide microarray

platforms used in genome-wide association studies (GWAS), which target common or less frequent SNPs with minor allele frequency (MAF)>0.01, and thus h^2_{SNP} generally represents the genetic contribution of these more common variants. Comparing h^2_{SNP} and h^2 (representing the entire proportion of a phenotype due to genetics) can inform the genetic architecture of a trait, representing the proportion due to common variants.

In this analysis, both narrow-sense and SNP-based heritability were estimated for fructosamine, glycated albumin and 1,5-AG using the same participants from the Atherosclerosis Risk in Communities (ARIC) Study, and compared across the different glycemic biomarkers.

2.3 Methods

Study population

The ARIC Study is a prospective cohort study initiated in 1987 to evaluate risk factors for cardiovascular disease in a community-based setting. Briefly, participants were recruited from four study sites: Forsyth, North Carolina; suburban Minneapolis, Minnesota; Jackson, Mississippi; and Washington County, Maryland. Overall, 15,792 middle-aged adults participated in the initial study visit (visit 1, 1987-1989), with 6 subsequent study visits (1990-2017). All study participants provided written informed consent, and the study protocols were approved by the relevant institutional review boards.[23]

Glycemic biomarkers

Samples for all glycemic biomarkers were collected at ARIC visit 2 (1990-1992). Fructosamine (Roche Diagnostics, Indianapolis IN, USA), glycated albumin (GA-L

Asashi Kasei Pharma Corporation, Tokyo, Japan) and 1,5-AG (GlycoMark, Winston-Salem, NC) were measured in 2012-2013 using a Roche Modular P800 system from samples stored at -70°C. Glucose was measured at visit 2 using the Roche Hitachi 911 analyzer using the hexokinase method (Roche Diagnostics). HbA1c was measured at visit 2 in stored whole blood samples using high performance liquid chromatography, using NGSP-certified assays standardized to the Diabetes Control and Complications Trial.[24]

Genotyping and Quality Control

Genotyping was performed using the Affymetrix 6.0 array. Samples with sex mismatches, genetic outliers, failed concordance with Taqman genotypes, or missingness >98% were excluded. First-degree relatives were defined by a DST value>0.8 (DST = IBS distance (IBS2 + 0.5*IBS1) / (N SNP pairs)) generated from PLINK.[25] Both members of each first-degree relative pair were included in the narrow-sense heritability estimation, and one member of a first-degree relative pair were excluded in SNP-based heritability estimation. SNPs were excluded if missingness was >5%, Hardy-Weinberg Equilibrium (HWE)<0.00001, or low minor allele frequency (MAF) <0.005. Imputation was pre-phased using ShapeIt (v1.r532) and then imputed using IMPUTE2 to 1,000 Genomes Phase I (March 2012).[26]

From the 30,038,522 imputed SNPs, SNPs were excluded if they had bases other than G, C, T or A, had duplicate base pair positions, imputation quality info score<0.99 and minor allele frequency (MAF)<0.01 to obtain a dataset with 3,224,517 SNPs. Imputed scores were converted to hard calls for the SNP-based heritability analyses using PLINK.[25]

Family-based study sample for narrow-sense heritability

Through genotyping, 384 first-degree relative pairs were identified (688 individuals, some were part of multiple pairs). Individuals were excluded if they met the following criteria: failed genetic quality control (N=50), did not attend visit 2 (N=29), did not fast for at least 8 hours (N=11) or missing fasting status (N=1), had diagnosed diabetes (N=32), or missing fasting glucose, HbA1c, fructosamine, glycated albumin, and 1,5-AG data (N=40). Individuals were further excluded if their related pair member did not pass quality control (N=91), potential parent-child relationships (N=20, as pedigree data was not available, types of first degree relatives could not be distinguished based on genetics alone and thus pairs with >15 year age difference were excluded), and likely monozygotic twins (first-degree relatives with the same age and sex, N=14), leaving 400 individuals who were members of sibling-pairs (**Supplemental Figure 2.1**).

Narrow-sense heritability analysis

Narrow-sense heritability was estimated using the variance components method using the program SOLAR-Eclipse.[27] This method uses a linear mixed model, with covariates (age, sex and ARIC study center) as fixed effects and genetics and environment as random effects. It partitions the variance between genetic and environmental effects and then heritability is calculated as the ratio of genetic variance to the total variance. The distributions of all glycemic biomarkers were skewed to the right among all participants (**Supplemental Figure 2.2**), and were therefore inverse normal transformed for all analyses.

SNP-based heritability study sample

The present study was restricted to self-identified white individuals because of limited power due to the smaller sample size of self-identified black participants (N=1,483 after exclusions; recommended sample size for h^2_{SNP}=4,000[24]). Of the 9,044 white ARIC participants with available genotyping data, participants with low quality genotype data (missingness>2%; N=290), did not attend ARIC visit 2 (N=313), did not fast for at least 8 hours (N=147) or missing fasting status (N=11), individuals with diagnosed diabetes (self-reported physician diagnosis or use of diabetes medications; N=480), or missing fasting glucose, HbA1c, fructosamine, glycated albumin, and 1,5-AG data (N=587), were excluded (**Supplemental Figure 2.1**).

LDAK SNP-based heritability analysis

The method Linkage Disequilibrium Adjusted Kinships (LDAK) was used to analyze SNP-based heritability for fasting glucose, HbA1c, fructosamine, glycated albumin and 1,5-AG.[21,28] This method employs a linear mixed model, with covariates such as age and sex as fixed effects and a genetic relationship matrix (GRM) calculated from genotyped SNPs for all pairs of individuals as random effects. The variance of the random effects is partitioned to isolate the variance due to genetics, and restriction maximum likelihood estimation is then used to estimate that variance. Heritability is then calculated as the proportion of total variance in the outcome due to genetics. The first step in LDAK is to calculate weights for each SNP, dividing the genome into approximately 1000kb sections and weighting SNPs based on the local LD structure such that areas of high LD had lower weights than those with low LD. The total weight of these SNPs was 113,120, representing the approximate number of independent loci evaluated.

Closely related individuals may affect SNP heritability analysis due to their shared environment or shared regions of LD, and hence we excluded them from analysis. To determine relatedness, kinship was calculated based on a thinned set of SNPs (not within 1Mb of each other or in LD, with $r^2>0.2$) using alpha = -0.25 (alpha is a parameter representing the relationship between heritability and MAF). Individuals were excluded so that no pair of individuals had a kinship value greater than the smallest observed kinship (-0.025, approximately no more related than cousins twice or thrice removed). Our analytic sample contained 5,575 individuals (**Supplemental Figure 2.1**). The ARIC study included a large percentage of married participants[29] (N=4,500 spousal pairs, 57% of individuals who attended visit 1), which represents a form of shared environment. However, the biomarker correlations among married couples was low (<0.10).

In each analysis, age, sex, ARIC study center, and the top 20 principal components (PC) were included as covariates. Predictor loadings from 1000 genomes were projected onto our data and the top 10 loadings were controlled for as recommended by the LDAK developers.[28] For each biomarker, strongly associated SNPs were evaluated using linear regression and inflation due to population substructure by calculating heritability separately in four chunks of chromosomes (chromosomes 1-3, 4-7, 8-11, 12-22) and comparing the sum of the heritabilities from the chunks to the heritability calculated using all of the chromosomes. If population substructure was present, the chromosomes would be correlated and hence the sum of heritability from the four chunks would be greater than heritability from all of the chromosomes (because each chunk would be representing more than just the heritability from the chromosomes in that chunk). Sensitivity analyses were performed excluding undiagnosed diabetes cases

(defined as fasting for at least 8 hours and glucose ≥126 mg/dL) from the heritability estimations.

GCTA SNP-based heritability analysis

Because there has been much debate but no consensus in the literature as to whether LDAK or the originally proposed SNP-based heritability method, Genome-wide Complex Trait Analysis (GCTA)[22] provide more accurate SNP-based heritability estimates, [28,30-32] both methods were used in this analysis. Due to sample size constraints, the most recent version of GCTA (GCTA-LDMS) did not run and therefore the original version of GCTA (GCTA-SC) was used. Individuals with kinship>0.05 were removed, leaving 6,443 individuals. A genetic relationship matrix was calculated and SNP-based heritabilities were estimated controlling for age, sex, ARIC study center and the first 10 principal components.

Bivariate heritability analyses

To explore the shared heritability among the glycemic biomarkers, bivariate heritability was performed, which calculates the percentage of heritability shared across two traits. Bivariate heritability models two traits as the outcome and estimates the genetic correlation between the traits. A negative correlation (between -1 and 0) indicates that the same genes increase the values of one trait while decreasing the values of the other trait, and a positive correlation (between 0 and 1) indicates that the same genes increase the values of both traits. Narrow-sense bivariate heritability was estimated using SOLAR-Eclipse and SNP-based bivariate heritability using GCTA.

2.4 Results

There were 5,575 unrelated individuals in the SNP-based heritability analytic sample and 400 first-degree relatives in the narrow-sense heritability analytic sample. Approximately half of the participants were female. Mean biomarker values were similar across both samples, and 5-7% of samples had undiagnosed diabetes (**Table 2.1**).

Narrow-sense heritability

The narrow-sense heritability estimates using sibling-pairs for 1,5-AG (h^2=0.55), glycated albumin (h^2=0.45) and fructosamine (h^2=0.44) were statistically significant ($p<1.9x10^{-4}$) (**Figure 2.1, Table 2.2**) and comparable to HbA1c (h^2= 0.34). The fasting glucose estimate was not significant (p=0.43), but analysis using visit 1 data (N=522) estimated heritability was 0.23 (p=0.03).

SNP-based heritability

The glycated albumin SNP-based heritability estimated using LDAK was (h^2_{SNP} = 0.30), followed by 1,5-AG (h^2_{SNP} = 0.17) and fructosamine (h^2_{SNP} = 0.13) (**Figure 2.1, Table 2.2**). HbA1c had similar SNP-based heritability to the nontraditional biomarkers (h^2_{SNP} = 0.30). The fasting glucose result was nonsignificant (p=0.11). Excluding undiagnosed diabetes cases (N=5,281) had little impact on these estimates (**Table 2.2**). Inflation for SNP-based heritability of the biomarkers was low (<3.3%).

SNP-based heritabilities estimated by GCTA were lower than estimates using LDAK. Fructosamine (h^2_{SNP} = 0.11), glycated albumin (h^2_{SNP} = 0.10), 1,5-AG (h^2_{SNP} = 0.15) and HbA1c (h^2_{SNP} = 0.17) (**Figure 2.1, Table 2.2**). The fasting glucose estimate was not significant (p=0.08).

Bivariate heritability analyses

Bivariate heritability estimates for fructosamine and glycated albumin showed shared genetics using SOLAR-Eclipse (0.46) with nearly complete overlap using GCTA (0.99). Glycated albumin and 1,5-AG had shared genetics which influence these traits in opposite directions, consistent with the inverse correlation of these biomarkers (-0.50 in SOLAR, -0.47 in GCTA); **Supplemental Table 2.1**). No other pairs of biomarkers had significantly shared heritability using GCTA or SOLAR-Eclipse.

2.5 Discussion

In this study, both narrow-sense and SNP-based heritabilities were estimated for nontraditional glycemic biomarkers. Because heritability is a population-specific measure that depends on relative genetic and environmental factors, it is important to estimate heritabilities in the same population in order to compare heritabilities across traits. This was done this for both traditional and nontraditional glycemic biomarkers, using the same population of white individuals participating in the ARIC Study.

Approximately half of the variation in fructosamine, glycated albumin and 1,5-AG was estimated to be controlled by genetic factors. Our results for 1,5-AG are consistent with previous estimations (0.55 in our study vs. 0.61 to 0.63 in previous studies).[11,20] There are no published reports of the heritability of fructosamine and glycated albumin. Our results illustrate that genetics play an important role in nontraditional glycemic biomarkers, and may affect these markers in a similar manner to HbA1c. Given that 60 variants are associated with HbA1c[33] (heritability = 0.20 to 0.75),[12,13,15-18] it is likely that more than the currently discovered (1 for fructosamine, 1 for glycated albumin, 7 for 1,5-AG)[34,35] variants are associated with nontraditional

glycemic biomarkers and these low numbers of SNPs may reflect the limited sample sizes to date for GWAS and sequencing studies for these biomarkers. The nonsignificant heritability estimates for fasting glucose may be in part due to its pre-analytic and intra-individual variability. Although nonsignificant, SNP-based heritability was similar to that estimated in a previous study in ARIC using participants from visit 1 (h^2_{SNP} =0.13 vs 0.8 in our LDAK analyses).[36] The significant and larger narrow-sense heritability estimated using visit 1 data (N=521, h^2=0.23, p=0.03) indicates that the smaller sample from visit 2 likely reduced power to estimate fasting glucose heritability.

As expected, the SNP-based heritabilities were lower than the narrow-sense heritabilities. SNP-based methods estimate heritability based on the variants for which data is collected (genotyped or sequenced), while narrow-sense heritability among related individuals is based on the entire genome. Additionally, narrow-sense heritability may be influenced by shared environment among family members, which does not impact SNP-heritability estimates. For all biomarkers, SNP-based heritabilities represented a portion of the narrow-sense heritability, indicating that the genetic influences on these traits is likely due to both common and rare variants.

Bivariate heritability was also estimated across all of the biomarkers, examining how much heritability is shared between them. The shared heritability between fructosamine and glycated albumin was significant, suggesting a substantial portion of overlapping genetics. However, this was expected due to the biological similarity of these two measures (80% of glycated proteins (i.e., fructosamine) are glycated albumin).[37,38] While the biomarkers in the present study all aim to capture hyperglycemia, some variability in these measures may be explained by non-glycemic

factors, which may be particularly important in the non-diabetic range. Alternatively, it could simply indicate the biomarkers are under control of different genes for other reasons such as the different time frames each biomarker represents (2-3 weeks for fructosamine and glycated albumin, 1-2 weeks for 1,5-AG 2-3 months for HbA1c, instantaneous for fasting glucose) or the differences in variability of these measures. Unfortunately, the lack of significance across the other bivariate analyses limit our ability to draw other conclusions about the amount of shared genetics across the other biomarkers.

The substantial heritability estimates for nontraditional biomarkers indicate a strong genetic component. Additional studies focused on the identification of genetic variants associated with fructosamine, glycated albumin and 1,5-AG will inform the biology of these biomarkers, and may identify limitations and implications for their clinical interpretation.

2.6 Tables and Figures

Table 2.1. Demographic and clinical characteristics in unrelated and first-degree relative study participants[1]

	SNP-based heritability	Narrow-sense heritability
	Unrelated Participants (N=5,575)	First-degree relatives (N=401)
Female	54%	55%
Age	57 (5.7)	58 (5.3)
Fructosamine (μmol/L)	227 (23)	227 (22)
Glycated albumin (%)	12.6 (1.6)	12.5 (1.5)
1,5-AG (μg/mL)	18.7 (5.7)	19.9 (6.4)
HbA1c (%)	5.4 (0.5)	5.5 (0.6)
Fasting glucose (mg/dL)	104 (17)	104 (16)
Undiagnosed diabetes[2]	5%	7%

[1]Continuous variables shown as mean (SD) and categorical variables shown as %
[2]Undiagnosed diabetes defined as fasting and glucose≥126 or nonfasting and glucose≥200

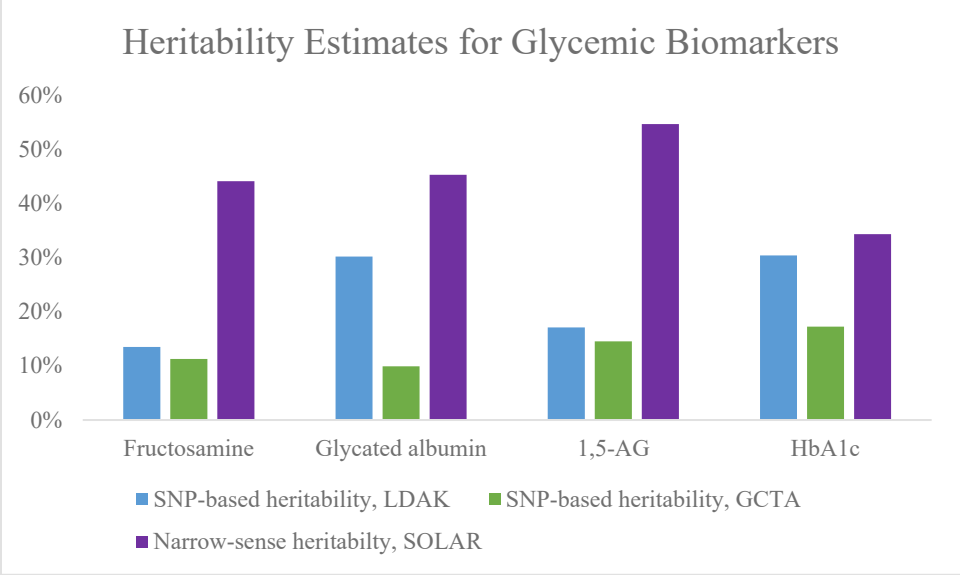
Table 2.2. SNP-based (h^2_{SNP})and narrow-sense (h^2) heritability estimates for glycemic biomarkers[1]

	h^2 **(SOLAR)**		h^2_{SNP} **(LDAK[2])**		h^2_{SNP} **(LDAK[2])** no undiagnosed diabetes		h^2_{SNP} **(GCTA)**	
	Heritability	P-value	Heritability (SD)	P-value	Heritability (SD)	P-value	Heritability (SE)	P-value
Fructosamine	0.44	2.9E-04	0.13 (0.06)	0.01	0.14 (0.07)	0.01	0.11 (0.05)	7.1E-03
Glycated albumin	0.45	1.8E-04	0.30 (0.07)	8.6E-07	0.27 (0.07)	1.7E-05	0.13 (0.04)	1.1E-03
1,5-AG	0.55	1.3E-05	0.17 (0.06)	3.6E-03	0.20 (0.07)	3.0E-03	0.11 (0.05)	7.2E-03
HbA1c	0.34	0.01	0.30 (0.07)	2.6E-06	0.37 (0.07)	1.1E-07	0.20 (0.05)	4.3E-06
Fasting glucose	2	0.43	8 (7)	0.11	13 (7)	0.03	5 (4)	0.13

[1]LDAK N=5575, relatedness cutoff=0.25; GCTA N=6443, relatedness cutoff=0.05; SOLAR N=400

[2]LDAK inflation <3.3% for all heritability estimates

Figure 2.1. Heritability estimates for glycemic biomarkers by method



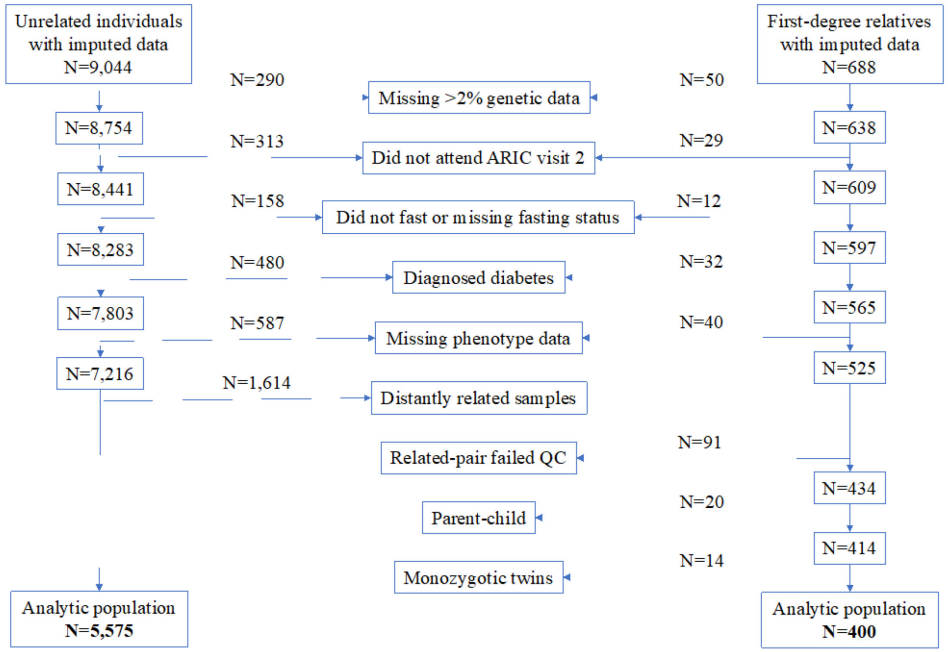
$p<0.05$ for all estimates

Supplemental Table 2.1. Bivariate heritability results

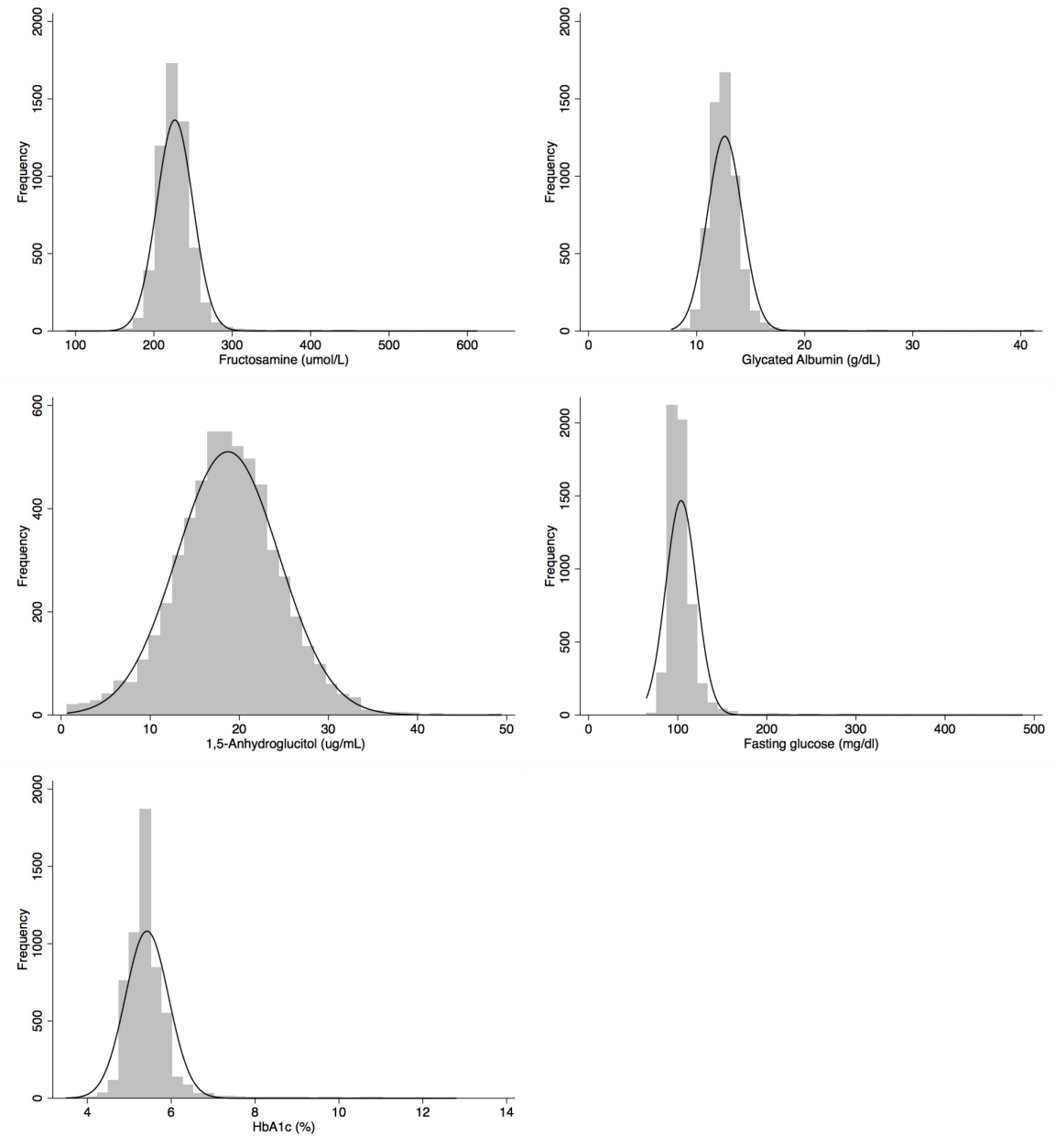
	Fructosamine	**Glycated Albumin**	**1,5-AG**	**HbA1c**	
Fructosamine		46 (p=0.04)*	-22 (P=0.25)	10 (p=0.70)	**SOLAR heritability (P-value)**
Glycated Albumin	100 (12)*		-50* (p=0.009)	3 (p=0.89)	
1,5-AG	-31 (21)	-47 (21)*		-30 (p=0.18)	
HbA1c	24 (19)	15 (20)	13 (18)		
GCTA heritability (SE)					

*p<0.05

Supplemental Figure 2.1. Study participant exclusions



Supplemental Figure 2.2. Untransformed biomarker distributions (N=5,575)



Chapter 3: Genome-wide association study of serum fructosamine and glycated albumin in adults without diagnosed diabetes: results from the Atherosclerosis Risk in Communities Study

Authors:

Stephanie J. Loomis, MPH*[1], Man Li, MS, PhD*[1,2], Nisa M. Maruthur, MD, MHS[1,3,4], Abigail S. Baldridge[5], Kari E. North, PhD[6], Hao Mei, MD, PhD[7], Alanna Morrison, PhD[8], April P. Carson, PhD, MSPH[5], James S. Pankow, PhD, MPH[9], Eric Boerwinkle, PhD[8], Robert Scharpf, PhD[10], Laura J. Rasmussen-Torvik, PhD[11], Josef Coresh, MD, PhD[1,3], Priya Duggal, PhD[1], Anna Köttgen, MD, MPH[1,12], Elizabeth Selvin, PhD, MPH[1,3]

*Authors contributed equally to manuscript.

[1]Department of Epidemiology, The Johns Hopkins University Bloomberg School of Public Health, Baltimore MD

[2]Division of Nephrology and Department of Human Genetics, University of Utah, Salt Lake City, Utah

[3]Welch Center for Prevention, Epidemiology, & Clinical Research, The Johns Hopkins University, Baltimore MD

[4]Division of General Internal Medicine, The Johns Hopkins University School of Medicine, Baltimore MD

[5]Department of Epidemiology, University of Alabama at Birmingham School of Public Health, Birmingham, AL

[6]Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC

[7]Department of Data Science, School of Population Health, University of Mississippi Medical Center, Jackson, MS

[8]Department of Epidemiology, The University of Texas Health Science Center at Houston School of Public Health at Houston, Houston, TX

[9]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN

[10]Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD

[11] Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL

[12]Institute of Genetic Epidemiology, Medical Center and Faculty of Medicine - University of Freiburg, Freiburg, Germany

3.1 Abstract

Fructosamine and glycated albumin are potentially useful alternatives to hemoglobin A1c (HbA1c) as diabetes biomarkers. The genetic determinants of fructosamine and glycated albumin, however, are unknown. We performed genome-wide association studies of fructosamine and glycated albumin among 2,104 black and 7,647 white participants without diabetes in the Atherosclerosis Risk in Communities (ARIC) Study and replicated findings in the Coronary Artery Risk Development in Young Adults (CARDIA) study. Among whites, rs34459162, a novel missense SNP in *RCN3,* was associated with fructosamine ($p=5.3x10^{-9}$), and rs1260236, a known diabetes-related missense mutation in *GCKR*, was associated with percent glycated albumin ($p=5.9x10^{-9}$) and replicated in CARDIA. We also found two novel associations among blacks: an intergenic SNP, rs2438321, associated with fructosamine ($p=6.2x10^{-9}$), and an intronic variant in *PRKCA,* rs59443763, associated with percent glycated albumin ($p=4.1x10^{-9}$), but these results did not replicate. Few established fasting glucose or HbA1c SNPs were also associated with fructosamine or glycated albumin. Overall, we found genetic variants associated with the glycemic information captured by fructosamine and glycated albumin, as well as with their non-glycemic component. This highlights the importance of examining the genetics of hyperglycemia biomarkers to understand the information they capture, including potential glucose-independent factors.

3.2 Introduction

Diabetes is defined by elevated blood glucose levels (hyperglycemia). Hemoglobin A1c (HbA1c) is formed as glucose binds to hemoglobin molecules within erythrocytes and is the standard clinical measure of chronic hyperglycemia used to diagnose and monitor diabetes(1). However, factors related to the nonglycemic portion of HbA1c such as erythrocyte turnover and hemoglobin characteristics can affect HbA1c values(2).

There is growing interest in fructosamine and glycated albumin, additional biomarkers of hyperglycemia that demonstrate associations with diabetes risk and complications similar to HbA1c(3-14). Fructosamine measures total serum protein bound to glucose. Glycated albumin is expressed as a percentage of serum albumin, the most abundant serum protein. Both biomarkers reflect glucose exposure over a shorter period of time (2-4 weeks) than HbA1c (2-3 months)(15). Fructosamine may be used to monitor glycemic control in clinical situations where HbA1c is problematic such as in the setting of anemia or hemoglobinopathies(16). While glycated albumin is not frequently used in the U.S., it is widely used in Japan and other countries as a complement to HbA1c to monitor short-term glycemic control(17).

The genetics of HbA1c and glucose have been well studied, however genetic factors that influence fructosamine and glycated albumin are uncharacterized. Of the known genome-wide association study (GWAS) variants associated with fasting glucose and HbA1c in European ancestry cohorts, few loci are associated with both(18). Many HbA1c variants are in genes related to hematologic factors rather than glucose metabolism,(19-21) while fasting glucose variants are in genes involved in glucose

metabolism (although these variants are not all associated with diabetes)(22-24). This lack of overlap suggests that some underlying genetic variants are specific to particular biomarkers of hyperglycemia rather than to type 2 diabetes. Understanding the genetic determinants of fructosamine and glycated albumin should help in the interpretation of these tests, and possibly extend our understanding of the pathophysiology of glucose metabolism. In particular, comparing the genetic overlap between different measures of glycemia may provide insight into the contributions of glycemic vs. nonglycemic gene variants, i.e. to what extent genetic factors operate via pathways directly relevant to diabetes pathophysiology ("glycemic") or operate via glycemic-independent pathways that do not influence glucose metabolism or diabetes risk ("nonglycemic", such as the hematologic variants associated with HbA1c). If nonglycemic genetic variants strongly impact fructosamine and glycated albumin, this may need to be taken into account in the interpretation of these biomarkers as measures of hyperglycemia.

We conducted GWAS of fructosamine and glycated albumin in blacks and whites in the Atherosclerosis Risk in Communities (ARIC) Study. We also compared previously identified genetic determinants for HbA1c and fasting glucose with fructosamine and glycated albumin to identify common genetic factors related to glucose metabolism and those that may be distinct to fructosamine and glycated albumin.

3.3 Methods

Study Population

The Atherosclerosis Risk in Communities (ARIC) Study is an ongoing, prospective cohort of 15,792 participants initiated in 1987 (25). Participants were middle-

aged adults recruited from four U.S. communities (Jackson, Mississippi; Forsyth County, North Carolina; Washington County, Maryland; and suburban Minneapolis, Minnesota). All study participants provided written informed consent and study protocols were approved by the relevant institutional review boards.

In the present study, we included 9,751 participants (7,647 whites and 2,104 blacks) who attended visit 2 (1990-1992), consented for use of DNA, did not have diagnosed diabetes (self-reported diagnosis or use of diabetes medications), had valid data on fructosamine and glycated albumin, and had genotyping data meeting quality control criteria (**Supplemental Figure 1**). Individuals with diagnosed diabetes were excluded to avoid potential bias caused by altered glucose levels as a result of diabetes treatment.

Genotyping

ARIC participants were genotyped using the Affymetrix 6.0 array and imputed separately by race using IMPUTE2(26) with the 1000 Genomes Project Phase I (March 2012) reference panel. Quality control excluded individuals based on SNP missing rate>5%, sex mismatch, high discordance with previous Taqman assay genotypes, genetic outlier status, and relatedness. SNPs with IMPUTE info score<0.8 or minor allele frequency (MAF)<0.05 were excluded. Only autosomal variants (on chromosomes 1-22) were considered. Principal components analysis was used to estimate population substructure with EIGENSTRAT(27).

Glycemic Markers

Fructosamine (Roche Diagnostics, Indianapolis IN, USA), glycated albumin and serum albumin (GA-L Asashi Kasei Pharma Corporation, Tokyo, Japan) were measured

in 2012-2013 using a Roche Modular P800 system from serum collected at visit 2 and stored at -70^{o}C(3). Percent glycated albumin was calculated per the manufacturer's protocol: [(glycated albumin concentration in g/dL / serum albumin concentration in g/dL)*100/1.14] + 2.9. We also examined total glycated albumin (g/dL) as well as serum albumin (g/dL) to help distinguish genetic factors specific to serum protein concentration vs hyperglycemia.

Serum glucose was measured on the Roche Hitachi 911 analyzer using the hexokinase method (Roche Diagnostics), and HbA1c was measured from whole blood stored at -70^{o}C using high performance liquid chromatography, standardized to the assay used in the Diabetes Control and Complications Trial(28).

Statistical Analysis

GWAS in blacks and whites were conducted using SNPTEST v2(29) for all glycemic biomarkers using imputed allele dosage and controlling for age, sex, field center, and the first 10 principal components under an additive genetic model. Fructosamine and glycated albumin (both percent (%) and total (g/dL)) were transformed on the natural log scale; therefore, the effect sizes are the change in the natural log of the biomarker per each additional risk allele. Exponentiating the effect sizes thus corresponds to the percent higher or lower biomarker levels per additional risk allele. Fasting glucose and HbA1c were not transformed. To identify additional independent SNPs associated with the traits, we performed conditional analyses for genome-wide significant finding. Using fructosamine, percent glycated albumin, and total glycated albumin as the dependent variable and the index SNP (SNP with the lowest p-value in a region showing a genome-wide significant association) as a covariate, we evaluated the

association between other SNPs with a minor allele frequency (MAF) ≥1% within 250 kB of the index SNP or between recombination hotspots surrounding the index SNP. To estimate percentage of variance explained by each SNP, we used the equation: $R_i^2 = b_i^2 * var(SNP_i)/var(y)$ where b_i = the effect size of the association between the SNP_i and the phenotype y, $var(SNP_i)$ is $2 * MAF_{SNPi} \times (1-MAF_{SNPi})$ and var(y) is the variance in the phenotype(30,31). We meta-analyzed across ancestries using a random-effects model by GWAMA(32).

In sensitivity analyses, we performed GWAS excluding undiagnosed diabetes cases (participants with fasting glucose ≥126 mg/dL or non-fasting glucose>200 mg/dl). To further evaluate the genetic variants pertaining to serum protein levels rather than hyperglycemia, we evaluated the top fructosamine and glycated albumin SNPs for association with total glycated albumin and with serum albumin (transformed on the natural log scale). To determine the extent to which glycemic biomarkers shared genetics, we calculated genetic correlations between the biomarkers using ARIC summary statistics with the LDSC program, using the precomputed LD scores from 1000 genomes European data(33).

Replication

Significant associations between genetic variants and fructosamine and percent glycated albumin in ARIC were evaluated for replication in the Coronary Artery Risk Development in Young Adults (CARDIA) cohort, a prospective cohort study initiated in 1985 to evaluate risk factors for heart disease among unrelated young adults(34).

Serum specimens from 2005-2006 were stored at –70°C and used to analyze glycated albumin and fructosamine in 2014 using a Roche COBAS 6000 chemistry

analyzer (Roche Diagnostics Corporation, Indianapolis, Indiana, USA). Glycated albumin and fructosamine were measured using the same assays used in ARIC (Lucica GA-L, Asahi Kasei Pharma Corporation, Tokyo, Japan; Roche Diagnostics, Indianapolis IN, USA; respectively).

Genotyping was performed using the Affymetrix 6.0 array. Standard quality control metrics were applied, and imputation to HapMap Phase II, Build 36, Release 22 was done using MACH(35). Genetic, covariate, fructosamine and glycated albumin data were available on 1,304 whites and 608 blacks. Individuals with diabetes (current use of glucose-lowering medications or fasting glucose ≥126 mg/dl were excluded from analysis. Linear regression analyses stratified by race were done for the association between significant ARIC SNPs and natural log transformed fructosamine and percent glycated albumin adjusted for age, sex, field center and the first three principal components.

Statistical analysis was done using SAS (v9.4, SAS Institute, Cary, NC) (data manipulation) and ProbABELv0.2(36). If the ARIC SNP was not available in the CARDIA dataset, we determined if proxy SNPs in linkage disequilibrium (LD) with the ARIC SNP ($r^2 > 0.7$ from 1000 genomes phase 3v5 European population, GRCh37 assembly using Single Nucleotide Polymorphisms Annotator (SNiPA)(37)) were available and if so, analyzed the associations with the proxy SNP. We considered a Bonferroni corrected one-sided p-value threshold of <0.05 (0.1/2 SNPs per race) for replication significance. We meta-analyzed ARIC and CARDIA results using fixed-effects inverse variance weighted model by METAL(38).

Candidate SNP analysis

We additionally evaluated previously-identified fasting glucose (n=41) and HbA1c (n=46) candidate SNPs from the NHGRI-EBI Catalog of published GWAS using the search terms "fasting glucose" and "HbA1c" (http://www.ebi.ac.uk/gwas/ as of 12/14/17) for association with fructosamine and percent glycated albumin in ARIC. SNPs were included if they were discovered in European ancestry cohorts, were genome-wide significant ($p<5x10^{-8}$) and were not in linkage disequilibrium (LD) with each other ($r^2 <$ 0.2, using SNiPA)(37). For the candidate SNP analyses, we used a study-wide significance threshold: two traits (fructosamine and glycated albumin), two races (black and white) and the number of candidate SNPs for each trait: $p<4.6x10^{-4}$ (0.05/(2*2*27)) for fasting glucose and $p<2.6x10^{-4}$ (0.05/(2*2*49)) for HbA1c.

We additionally performed analyses controlling for these known fasting glucose and HbA1c SNPs compiled into scores. Scores were calculated as the sum of the number of risk alleles, weighted by the effect size in ARIC among whites.

Comparison of variance explained

To compare the influence of glycemic and nonglycemic genetic variants on fructosamine and glycated albumin to that of HbA1c, we calculated the percent phenotypic variance explained by the SNPs in our study to those from published results of HbA1c. In addition, we calculated variance explained by known fasting glucose and albumin SNPs (identified by the same criteria used for fasting glucose and HbA1c). Percent variance explained was calculated using the equation described above.

3.4 Results

Overall, 7,647 whites and 2,104 blacks from ARIC and 1,304 whites and 608 blacks from CARDIA were included in this study (**Supplemental Table 3.1**). Mean age in ARIC was higher (56-57 years), than in CARDIA (45-46 years). The cohorts had similar distribution of sex. CARDIA had a greater percentage of black participants and lower mean values for each measure of glycemia as compared to ARIC.

We identified four genome-wide significant loci in ARIC, two associated with fructosamine and two with percent glycated albumin. Three of these variants, rs34459162 intronic to *RCN3*, rs2438321 (intergenic), and rs59443763 intronic to *PRKCA*, have not previously been reported to be associated with any glycemic traits in humans (**Table 3.1**). None of the analyses showed evidence for inflation (**Supplemental Figures 3.2-3.7**).

Among whites, rs34459162 (MAF=0.08), a missense SNP in *RCN3* on chromosome 19, was significantly associated with 1.8% lower fructosamine per minor allele ($p=5.3x10^{-9}$, variance explained=0.6%; **Table 3.1, Figures 3.1-3.2**). This SNP was also associated with total glycated albumin ($p=3.8x10^{-8}$; **Table 3.2)**. The association with percent glycated albumin approached genome-wide significance ($p=7.3x10^{-8}$), but this SNP was not associated with fasting glucose or HbA1c (**Table 3.2**). A proxy for rs34459162, (rs8105626, in $r^2=1$ rs34459162) was not associated in CARDIA for association with fructosamine (p=0.09) although the effect sizes were identical and meta-analysis across the cohorts was significant ($p=4.9x10^{-9}$, **Table 3.1**). Conditional analysis in ARIC showed that the additional 63 significant SNPs in the region became nonsignificant after conditioning on rs34459162. In blacks, rs34459162 did not meet the info score>0.8 threshold and thus was not analyzed (**Table 3.2)**.

Among whites, rs1260326 (also known as rs343480), a known missense mutation in *GCKR* on chromosome 2 (MAF=0.41), was significantly associated with 1.1% lower levels of percent glycated albumin per minor allele ($p=5.3x10^{-9}$, variance explained=0.3%; **Supplemental Figure 3.8-3.9**, **Table 3.1**). The association with percent glycated albumin was also significant in CARDIA (p=0.04), with similar percent difference (0.8% lower per minor allele) and genome-wide significant meta-analysis results ($2.3x10^{-8}$, **Table 3.1**). The conditional analysis did not reveal additional independent signals in this region. This SNP was not associated with any biomarker among blacks (MAF=0.14, **Table 3.2, Supplemental Table 3.2**), but power was limited, and the meta-analysis across ancestries was not significant (**Table 3.1, Supplemental Figure 3.10**).

Among blacks, rs2438321 on chromosome 11 (MAF=0.11) was associated with 3.5% higher levels of fructosamine per minor allele at a genome-wide significant level ($p=6.2x10^{-9}$, variance explained=1.8%; **Table 3.1, Supplemental Figure 3.11-3.12**), and approached significance with percent glycated albumin ($p=6.4x10^{-5}$), and total glycated albumin ($p=2.0x10^{-6}$) (**Table 3.2**). Rs2438321 was not associated with HbA1c in blacks and was not associated with any of the markers of hyperglycemia in whites independently or in a meta-analysis across ancestries (**Table 3.1**, **Table 3.2**). This SNP was not available in the CARDIA dataset, however a proxy SNP in perfect LD ($r^2=1$, rs35256014, MAF=0.06) was present but did not replicate the association with fructosamine in blacks (p=0.57) and meta-analysis were also nonsignificant (**Table 3.1, Supplemental Figure 3.13**).

An intronic variant, rs59443763, in *PRKCA* on chromosome 17 (MAF=0.06), was significantly associated with 5.4% higher percent glycated albumin per minor allele in blacks ($p=4.9x10^{-9}$), variance explained=2%; **Table 3.1**, **Supplemental Figure 3.14-3.15**). It was also associated with fructosamine ($p=9.4x10^{-7}$) and total glycated albumin ($p=5.8x10^{-7}$) although these associations did not meet genome-wide significance (**Table 3.2**). This SNP was not significant among whites nor in trans-ancestry meta-analysis (**Table 3.1, Table 3.2**), but there was limited power to replicate (**Supplemental Table 3.2**). No proxy SNPs with $r^2>0.7$ were available in the CARDIA dataset and thus replication was not possible for this association.

Sensitivity analyses

In analyses which excluded participants with undiagnosed diabetes, genome-wide significant results remained for the white sample (N=7,229), but were no longer present among the reduced sample of black participants (N=1,878) (**Supplemental Table 3.3**).

Of the SNPs significantly associated with fructosamine or glycated albumin, only rs2438321 in blacks (p=0.002) was significantly associated with serum albumin with a Bonferroni corrected p-value (0.05/(4 SNPs * 2 races) = 0.006) (**Supplemental Table 3.4**).

Controlling for fasting glucose or HbA1c variants did not reveal any additional genome-wide significant variants, but for glycated albumin controlling for HbA1c increased significance of rs34459162 ($p=3.77x10^{-8}$) and controlling for fasting glucose score attenuated the p value for rs1260326 (p=0.0006) among whites.

Fructosamine, percent glycated albumin and total glycated albumin had strong, statistically significant genetic correlations (0.92 to 1.17) indicating a large proportion of

shared genetics (**Supplemental Table 3.5**). Correlations between fasting glucose and HbA1c with the other biomarkers were moderate to substantial but were not significant.

Candidate SNP analysis

We investigated SNPs previously identified in fasting glucose and HbA1c GWAS for association with fructosamine and percent glycated albumin. Nineteen of the 41 fasting glucose SNPs were nominally ($p<0.05$) associated with fasting glucose, and 13 of these associations were in the same direction in blacks and whites (**Table 3.3, Supplemental Table 3.6**) and all but 8 were in the same direction as the discovery cohort in whites. Four variants (10%) were study-wide significantly associated with fructosamine and percent glycated albumin in whites, three of which were associated with percent glycated albumin and one of which was associated with fructosamine. No variants were study-wide significantly associated with fructosamine or percent glycated albumin in blacks.

Thirty-one of 46 previously-identified HbA1c SNPs were nominally associated with HbA1c, 15 of which were associated in the same direction in blacks and whites (**Table 3.4, Supplemental Table 3.7**) and all but three in the same direction as the discovery cohort in whites. Five SNPs (11%) demonstrated a study-wide significant association with fructosamine or glycated albumin in whites. All variants associated with multiple glycemic biomarkers had effects in the same direction.

Percent Variance Explained

SNPs associated with fasting glucose (N=41, listed in Table 3) explained 1.4% of the variance in fructosamine, 3.2% of the variance in percent glycated albumin and 1.9% of the variance in total glycated albumin among whites. Taking SNPs associated with

serum albumin from the GWAS catalogue explained 0.4% of the variance of fructosamine, 1.1% of the variance of percent glycated albumin and 0.7% of the variance of total glycated albumin among the white sample.

3.5 Discussion

We identified four SNPs significantly associated with fructosamine and glycated albumin among either whites or blacks, one which replicated in a second cohort, and three not previously associated with glycemic traits. Several known fasting glucose and HbA1c SNPs were significantly associated with fructosamine or glycated albumin.

Among whites, rs1260326 was associated with percent glycated albumin. This variant reflects the same signal associated with type 2 diabetes and fasting glucose: it is in perfect LD with a known type 2 diabetes variant (r^2=1 among 1000 genomes phase 3 Europeans with rs145819220, from a recent large type 2 diabetes GWAS(39)) and in strong LD with a known fasting glucose variant (r^2=0.91 with rs780094(38,39). Rs1260326 is located in glucokinase (hexokinase 4) regulator (*GCKR*), which encodes a regulatory protein primarily active in the liver that inhibits glucokinase (GCK), the enzyme in the first step of glycolysis and involved in converting glucose to glycogen for storage. GCK is considered a glucose sensor that helps maintain glucose homeostasis. The *GCKR* protein product inhibits the activity of GCK, increasing serum glucose levels. *GCKR* is an established type 2 diabetes gene(40-42), and is associated with multiple other traits including kidney disease, triglyceride levels and Crohn's disease(43-45). Thus, this variant likely represents part of a glycemic pathway, but it is interesting that in our study it is only significantly associated with one measure of hyperglycemia, approached

significance with fasting glucose (although controlling for fasting glucose score made this variant nonsignificant) and total glycated albumin, but is not associated with fructosamine or HbA1c, given the moderate to strong correlations and genetic correlations among the biomarkers (**Supplemental Table 3.8, 3.5**). That *GCKR* is primarily expressed in the liver rather than the pancreas(46,47) aligns with the finding of association with fasting glucose, which measures hepatic glucose output. Albumin is also produced by the liver, while erythrocytes and hemoglobin are not likely affected by liver function, thus perhaps hepatic-specific genetic factors would be more likely to associate with percent glycated albumin levels than with HbA1c. It is also possible that HbA1c, affected by other glucose-altering factors, may mask the effect of rs1260326 on *GCKR*. Adjustment for serum albumin may explain the association with percent glycated albumin but not fructosamine.

We also identified several variants of potential interest which were significant in ARIC but lacked replication. Among whites, rs34459162, in *RCN3*, was associated with fructosamine and total glycated albumin. *RCN3* encodes reticulocalbin 3, an EF-hand calcium binding domain(48). This SNP was not associated with serum albumin in our analysis, but a SNP in perfect LD with rs34459162, rs2280401, was associated with total protein in a Japanese population(49) and serum albumin in an East Asian population(50), indicating a possible impact on fructosamine and glycated albumin through nonglycemic pathways. Among blacks, we found two novel variants: rs2438321 (intergenic and closest to *CNTN5* which encodes a glycosylphosphatidylinositol (GPI)-anchored neuronal membrane protein, a member of the immunoglobulin superfamily and the contactin family.) associated with fructosamine and rs59443763 (*PRKCA* which encodes protein

kinase C alpha, ubiquitous in cellular processes) associated with percent glycated albumin at a genome-wide level of significance. There is no prior literature on either of these as potential glycemic loci in diabetes. While we had sufficient power to replicate the results for whites in CARDIA (**Supplemental Table 3.2**), we had low power among blacks, which may be why these SNPs did not replicate. These variants became nonsignificant after excluding undiagnosed diabetes, which may be due to the greater number of individuals with undiagnosed diabetes among blacks than whites. Blacks had higher values of glycemic biomarkers, thus removing undiagnosed diabetes cases could have had a greater impact on associations among blacks than whites. These variants should be evaluated in larger African ancestry datasets as they become available. Rs34459162, rs2438321 and rs59443763 are of potential interest, but as these SNPs currently lack replication, we cannot rule out false positive results.

Results varied by ancestry for the SNPs available in both blacks and whites: neither SNP was significant in both blacks and whites, and meta-analyses were nonsignificant. While this may partially be explained by differing allele frequencies (rs1260326: 0.41 in whites, 0.14 in blacks; rs2438321: 0.24 in whites, 0.11 in blacks), a differential effect by ancestry on fructosamine and glycated albumin is also possible. This may be particularly true for rs2438321, where the direction of effect differs across ancestries.

In addition to investigating fructosamine and glycated albumin individually, comparing to traditional glycemic markers (fasting glucose and HbA1c) can help to clarify the biological pathways involved in diabetes. Fasting glucose-related SNPs explained almost twice the variance of percent glycated albumin than that of

fructosamine. This may reflect the adjustment for serum albumin with percent glycated albumin and not with fructosamine, allowing percent glycated albumin levels to be influenced more by glucose levels and less by albumin levels. However, albumin SNPs also explained more variance of percent glycated albumin than that of fructosamine or total glycated albumin. Given the small percentages, it is difficult to draw first conclusions from these results.

Only five HbA1c variants were significantly associated with fructosamine or glycated albumin. This is consistent with the findings that the majority of HbA1c variants are related to erythrocyte and hemoglobin factors that we would not expect to be related to fructosamine or glycated albumin. Many associations of fructosamine or glycated albumin with HbA1c SNPs or fasting glucose SNPs were present in whites but not blacks. This is not surprising given the SNPs were originally detected in whites, and our sample size was larger for whites, with corresponding higher power to detect moderate associations. Not all of the previously discovered SNPs for fasting glucose and HbA1c replicated for those outcomes in our sample, but this again may have to do with lack of power.

We found both glycemic and nonglycemic genetic factors influenced fructosamine and glycated albumin levels. We identified a likely glycemic variant in a gene associated with type 2 diabetes (*GCKR*) supporting its role in diabetes biology, and a likely nonglycemic variant in a gene (*RCN3*) that may reflect the biology of a biomarker (i.e., influencing amount of serum protein available to be glycated) rather than the biology of type 2 diabetes. This contribution of glycemic and nonglycemic variants is similar to the pattern of genetic contribution to HbA1c, for which the majority of genetic

variants are nonglycemic(19,20). In our study, previously identified nonglycemic variants(19-21) explained 3.4% of the variance in HbA1c, and the glycemic variants explained 2.1% (**Supplemental Table 3.9**). Despite the previous studies having much larger sample sizes (and thus more power to detect associations with HbA1c), the percent variance explained we found for fructosamine (0.6% by likely nonglycemic rs34459162) and glycated albumin (0.3% by likely glycemic rs1260326) was of a similar magnitude. Both Soranzo and Chen found that taking nonglycemic variants into account modestly impacted diabetes reclassification, and Wheeler found a more substantial effect(19-21). Given that that future, larger studies on fructosamine and glycated albumin will likely reveal other significant variants, it will be important to determine if the effect of nonglycemic variants is substantial enough to impact the clinical interpretation of fructosamine and glycated albumin.

A major limitation of this study was the limited sample size, particularly the smaller sample size in blacks. The differences in ancestries make replication of results difficult, particularly if allele frequencies differ, and warrant more studies focused on multi-ethnic populations. Also, the lack of an available SNP or proxy for rs59443763 in CARDIA, possibly due to the imputation reference panel (HapMap Phase II), impeded our ability to evaluate replication of this finding. In addition, the sample size for our replication cohort was much smaller than our discovery cohort, limiting our power to replicate the significant ARIC findings in blacks.

In summary, through GWAS in a community-based population of blacks and whites we identified and replicated two significant variants associated with fructosamine and/or glycated albumin, one of which was novel. These variants map into a likely

glycemic, known diabetes gene, and a likely nonglycemic gene. This highlights the utility of examining genetics of diabetes biomarkers both for providing insight into the pathophysiology of diabetes and for better understanding glucose–independent influences on measures of hyperglycemia.

3.6 Tables and Figures

Table 3.1. Genome-wide significant loci for fructosamine and percent glycated albumin.

					ARIC*					CARDIA*				ARIC+CARDIA*			ARIC White+Black*		
SNP‖	**Gene**	**A1/ A2†**	**Outcome**	**Race**	**A1 freq**	**Beta‡ (SE)**	**% diff§**	**P-value**	**% var expl**	**A1 freq**	**Beta‡ (SE)**	**% diff§**	**P-value**	**Beta‡ (SE)**	**% diff§**	**P-value**	**Beta‡ (SE)**	**% diff§**	**P-value**
rs34459162	*RCN3*	C/T	Fructosamine (umol/L)	W	0.08	-0.02 (0.003)	-2%	5.3E-9	0.6%	0.09	-0.02 (0.014)	-2%	0.09	-0.02 (0.003)	-2%	4.9E-9	-	-	-
rs1260326	*GCKR*	T/C	Percent glycated albumin (%)	W	0.41	-0.01 (0.002)	-1%	5.9E-9	0.3%	0.43	-0.01 (0.004)	-1%	0.04	-0.01 (0.002)	-1%	2.3E-8	-0.007 (0.005)	- 0.7%	0.14
rs2438321	*CNTN5*	G/A	Fructosamine (umol/L)	B	0.11	0.03 (0.006)	3%	6.2E-9	1.8%	0.06	0.006 (0.011)	0.6%	0.57	0.03 (0.005)	3%	2.9E-8	0.02 (0.02)	2%	0.37
rs59443763	*PRKCA*	C/T	Percent glycated albumin (%)	B	0.06	0.05 (0.009)	5%	4.1E-9	2.0%	-	-	-	-	-	-	-	-	-	-

*ARIC: N=7,647 whites, 2,104 blacks; CARDIA: N=1,304 whites, 608 blacks; ARIC+CARDIA is a meta-analysis across the cohorts, ARIC White+Black is a meta-analysis across the ancestries in ARIC

†A1 is the minor allele in whites

‡Mean change in ln(outcome) for each additional A1 allele

§Percent higher or lower levels of the outcome for each additional copy of the minor allele, calculated as e^{β} *100

‖rs34459162 and rs2438321 not available in CARDIA dataset, evaluated proxy SNPs rs8105626 and rs35256014, respectively in perfect LD (r^2=1 with ARIC SNP).

Table 3.2. Genome-wide significant loci for fructosamine and percent glycated albumin and their association with total glycated albumin, fasting glucose and HbA1c in ARIC*,†

SNP	Closest Gene	A1/ A2‡	Race	1000 genomes A1 freq§	A1 freq	Fructosamine (umol/L)		Percent glycated albumin (%)		Total glycated albumin (g/dL)		Fasting glucose (mg/dL)		HbA1c (%)	
						Beta (SE)	P-value	Beta (SE)	P-value	Beta (SE)	P-value	Beta (SE)	P-value	Beta (SE)	P-value
rs34459162	*RCN3*	C/T	White	0.08	0.08	**-0.02 (0.003)**	**5.3E-09**	-0.02 (0.004)	7.3E-08	**-0.03 (0.005)**	**3.8E-08**	0.26 (0.56)	0.64	-0.02 (0.02)	0.25
			Black	-	-	-	-	-	-	-	-	-	-	-	-
rs1260326	*GCKR*	T/C	White	0.43	0.41	-0.003 (0.002)	0.02	**-0.01 (0.002)**	**5.9E-09**	-0.01 (0.003)	1.3E-05	-1.13 (0.28)	4.5E-05	-0.01 (0.008)	0.09
			Black	0.10	0.14	0.0005 (0.005)	0.93	-0.0003 (0.006)	0.96	0.001 (0.009)	0.9	-0.53 (1.16)	0.64	-0.01 (0.04)	0.76
rs2438321	*CNTN5*	G/A	White	0.30	0.24	-0.001 (0.002)	0.44	0.0005 (0.002)	0.82	-0.001 (0.003)	0.71	-0.21 (0.32)	0.52	0.001 (0.01)	0.95
			Black	0.11	0.11	**0.03 (0.006)**	**6.2E-09**	0.03 (0.007)	6.4E-05	0.05 (0.009)	2.0E-06	4.19 (1.24)	8.3E-04	0.14 (0.04)	6.8E-04
rs59443763	*PRKCA*	C/T	White	0.005	-	-	-	-	-	-	-	-	-	-	-
			Black	0.09	0.06	0.04 (0.008)	9.4E-07	**0.05 (0.009)**	**4.1E-09**	0.06 (0.01)	5.8E-07	7.2 (1.65)	9.6E-06	0.18 (0.05)	8.3E-04

*Genome-wide significant results are in bold

†Fructosamine, glycated albumin percent and total glycated albumin are log transformed

‡A1 is the minor allele in whites

§1000 genomes populations: White=CEU, Black=ASW

Table 3.3. Significance of associations between fasting glucose known genetic determinants and fructosamine and percent glycated albumin in ARIC*,†

			White				Blacks			
SNP	**Closest Gene**	**A1/ A2**	**A1 freq**	**P-value fructos-amine**	**P-value glycated albumin**	**P-value fasting glucose**	**A1 freq**	**P-value fructos-amine**	**P-value glycated albumin**	**P-value fasting glucose**
rs10747083	P2RX2	G/A	0.33	0.95	0.76	0.14	0.16	0.17	0.11	0.95
rs10830963	MTNR1B	G/C	0.28	0.001	0.003	**3.3E-07**	0.07	0.13	0.03	0.38
rs10885122	ADRA2A	T/G	0.12	0.82	0.75	0.76	0.68	0.24	0.72	0.21
rs11071657	C2CD4B	G/A	0.38	0.50	0.81	0.52	0.12	0.98	0.95	0.43
rs11558471	SLC30A8	G/A	0.32	0.005	**8.2E-05**	**1.8E-05**	0.09	0.31	0.18	0.65
rs11603334	ARAP1	A/G	0.16	0.01	0.08	0.47	0.06	0.79	0.95	0.76
rs11605924	CRY2	A/C	0.47	0.36	0.58	0.66	0.84	0.58	0.72	0.44
rs11708067	ADCY5	G/A	0.23	0.07	0.001	**0.01**	0.16	0.002	0.003	**0.03**
rs11715915	AMT	T/C	0.30	0.30	0.74	**0.04**	0.23	0.06	0.37	0.31
rs11920090	SLC2A2	A/T	0.13	0.004	0.002	0.34	0.35	0.47	0.30	0.28
rs13179048	PCSK1	A/C	0.31	0.11	0.13	0.51	0.08	0.18	0.52	0.78
rs1371614	DPYSL5	T/C	0.25	0.47	0.78	0.69	0.36	0.18	0.11	0.21
rs143399767	YRNA	C/A	0.01	0.65	0.24	0.55				
rs1483121	OR4S1	A/G	0.13	0.41	0.36	0.74	0.04	0.46	0.42	0.94
rs16913693	IKBKAP	G/T	0.03	0.60	0.78	0.11	0.25	0.73	0.68	0.91
rs174550	FADS1	C/T	0.33	0.16	0.28	0.54	0.08	0.37	0.15	0.93
rs17762454	RREB1	T/C	0.26	0.13	0.32	**0.001**	0.16	0.16	0.76	0.75
rs2191349	TMEM195, DGKB	G/T	0.46	0.17	0.36	**0.02**	0.44	0.58	0.73	**0.004**
rs2293941	PDX1	A/G	0.22	0.02	0.005	**0.03**	0.17	0.38	0.60	0.77
rs2302593	GIPR	G/C	0.50	1.00	0.40	**0.02**	0.30	0.64	0.94	0.97
rs2657879	GLS2	G/A	0.18	0.93	0.94	0.31	0.06	0.05	0.23	**0.04**
rs2722425	ZMAT4	T/C	0.12	0.01	0.01	0.78	0.37	0.25	0.35	0.41
rs340874	PROX1	T/C	0.45	0.42	0.82	0.40	0.17	0.36	0.74	0.15
rs35767	IGF1	G/A	0.16	0.76	0.54	0.62	0.55	0.43	0.10	**0.02**
rs3736594	MRPL33	C/A	0.26	0.11	0.07	0.25	0.43	0.07	0.03	0.62
rs3783347	WARS	T/G	0.22	0.45	0.51	0.57	0.06	0.44	0.60	0.94
rs3829109	DNLZ	A/G	0.29	0.003	0.13	0.16	0.18	0.51	0.39	0.46
rs4506565	TCF7L2	T/A	0.31	3.9E-04	**2.5E-05**	**2.7E-05**	0.44	0.09	0.40	0.83
rs4607517	GCK	A/G	0.17	**2.0E-04**	0.003	**5.0E-04**	0.11	0.09	0.10	0.08
rs4841132	PPP1R3B	A/G	0.09	0.33	0.32	**0.05**	0.13	0.74	0.29	**0.05**
rs560887	G6PC2	T/C	0.30	0.08	0.003	**7.3E-05**	0.05	0.81	0.58	0.91
rs576674	KL	G/A	0.16	0.45	0.50	0.33	0.61	0.17	0.26	0.11
rs6048205	FOXA2	G/A	0.05	0.79	0.95	0.07	0.18	0.02	0.25	**0.05**
rs6072275	TOP1	A/G	0.16	0.57	0.30	**0.01**	0.08	0.13	0.12	0.08
rs6943153	GRB10	T/C	0.31	0.67	0.46	0.16	0.70	0.09	0.77	0.61
rs7034200	GLIS3	A/C	0.49	0.19	0.04	**0.03**	0.61	0.18	0.02	0.11
rs7651090	IGF2BP2	G/A	0.31	0.01	0.03	**0.003**	0.56	0.32	0.58	0.40
rs7708285	ZBED3	G/A	0.30	0.32	0.64	0.71	0.15	0.56	0.52	0.16
rs780094	GCKR	T/C	0.40	0.05	**5.7E-05**	**1.0E-04**	0.18	0.75	0.96	0.90
rs7944584	MADD	T/A	0.27	0.56	0.76	**0.05**	0.04	0.98	0.38	0.77
rs9368222	CDKAL1	A/C	0.27	0.19	0.09	0.27	0.20	0.96	0.80	0.68

*Candidate SNPs selected from NHGRI database based on previous genome-wide significant associations

†P-values that reach study-wide significance for fructosamine and glycated albumin, $p<3.0x10^{-4}$ (0.05/(2*2*41)) are in bold; P values that reach nominal significance ($p<0.05$) for fasting glucose in bold

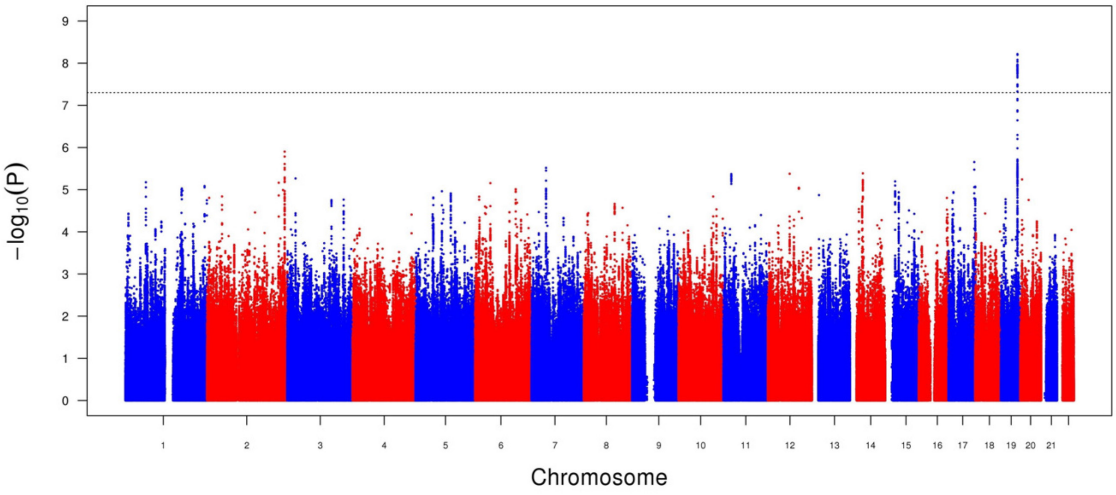
Table 3.4. Significance of associations between HbA1c known genetic determinants and fructosamine and percent glycated albumin in ARIC*,†

			Whites				Blacks			
SNP	**Closest Gene**	**A1/ A2**	**A1 freq**	**P-value fructos-amine**	**P-value glycated albumin**	**P-value HbA1c**	**A1 freq**	**P-value fructos-amine**	**P-value glycated albumin**	**P-value HbA1c**
rs1046896	FN3KRP	T/C	0.31	0.034	0.048	0.052	0.24	0.661	0.455	0.585
rs10774625	ATXN2	G/A	0.48	0.571	0.032	**0.008**	0.09	0.221	0.142	0.503
rs10823343	HK1	G/A	0.26	0.452	0.065	**1.1E-06**	0.45	0.808	0.529	0.899
rs10830963	MTNR1B	G/C	0.28	0.001	0.003	**0.006**	0.07	0.134	0.035	0.222
rs11248914	ITFG3	C/T	0.35	0.011	0.229	0.055	0.36	0.067	0.354	0.485
rs11558471	SLC30A8	G/A	0.32	0.005	**8.2E-05**	**0.010**	0.09	0.309	0.176	0.989
rs11603334	ARAP1	A/G	0.16	0.005	0.078	0.228	0.06	0.788	0.950	0.734
rs11708067	ADCY5	G/A	0.23	0.070	0.001	**0.034**	0.16	0.002	0.003	0.074
rs11954649	SOX30	G/C	0.00	NA	NA	NA	0.05	0.812	0.220	0.184
rs11964178	C6orf183	G/A	0.43	0.861	0.844	0.702	0.36	0.180	0.748	0.475
rs12621844	FOXN2	C/T	0.39	0.732	0.294	0.358	0.84	0.829	0.853	0.666
rs12819124	RP1	A/C	0.47	0.765	0.930	**0.001**	0.21	0.500	0.826	0.146
rs13134327	FREM3	A/G	0.32	0.800	0.662	**0.031**	0.30	0.220	0.854	0.772
rs1402837	G6PC2	T/C	0.22	0.016	**9.3E-05**	**2.7E-07**	0.31	0.471	0.728	0.656
rs1558902	FTO	A/T	0.41	0.545	0.017	0.131	0.11	0.065	0.176	0.514
rs17509001	ATAD2B	C/T	0.14	0.309	0.152	**0.027**	0.09	0.604	0.524	0.802
rs17533903	MYO9B	A/G	0.21	0.577	0.167	**5.0E-04**	0.24	0.403	0.232	0.689
rs17747324	TCF7L2	C/T	0.23	**1.7E-04**	**3.2E-05**	**1.8E-04**	0.07	0.396	0.755	0.890
rs1800562	HFE	A/G	0.06	0.337	0.128	**0.001**	0.01	0.846	0.670	0.146
rs198846	HFE	A/G	0.16	0.390	0.198	0.123	0.88	0.533	0.270	0.260
rs2110073	PHB2	T/C	0.10	0.321	0.938	**0.002**	0.42	0.177	0.349	0.160
rs2383208	MTAP	G/A	0.18	0.008	0.019	**0.007**	0.19	0.293	0.763	0.596
rs2408955	SENP1	G/T	0.48	0.567	0.879	**0.002**	0.63	0.813	0.824	0.632
rs267738	CERS2	G/T	0.21	0.444	0.434	0.511	0.04	0.853	0.745	0.297
rs2779116	SPTA1	T/C	0.27	0.714	0.792	**9.9E-06**	0.22	0.111	0.167	0.961
rs282587	ATP11A	G/A	0.12	0.586	0.248	**1.0E-04**	0.69	0.050	0.142	0.246
rs3824065	GCK	T/C	0.42	3.3E-04	0.002	0.125	0.24	0.086	0.134	0.910
rs4607517	GCK	A/G	0.17	**2.0E-04**	0.003	**5.7E-05**	0.11	0.088	0.097	0.244
rs4737009	ANK1	A/G	0.24	0.126	0.515	**0.018**	0.46	0.274	0.461	0.361
rs4745982	HK1	G/T	0.07	0.190	0.263	**5.8E-06**	0.10	0.889	0.921	0.602
rs4820268	TMPRSS6	G/A	0.46	0.264	0.350	**0.010**	0.72	0.073	0.029	0.181
rs560887	G6PC2	T/C	0.30	0.080	0.003	**0.003**	0.95	0.811	0.583	0.861
rs579459	ABO	C/T	0.23	0.268	0.340	**0.001**	0.13	0.210	0.487	0.600
rs592423	CITED2	A/C	0.46	0.514	0.535	0.063	0.61	0.323	0.605	0.333
rs6474359	ANK1	C/T	0.03	0.979	0.486	**4.3E-04**	0.27	0.077	0.047	0.879
rs6980507	SLC20A2	A/G	0.39	0.256	0.873	**0.015**	0.48	0.514	0.557	0.663
rs7040409	C9orf47	G/C	0.07	0.499	0.765	**0.001**	0.26	0.028	0.005	0.115
rs7616006	SYN2	G/A	0.43	0.705	0.385	0.181	0.36	0.974	0.959	0.836
rs761772	TMC6	C/T	0.13	0.077	0.109	**0.009**	0.13	0.480	0.361	0.861
rs7756992	CDKAL1	G/A	0.27	0.180	0.094	0.292	0.42	0.856	0.997	0.995
rs8192675	SLC2A2	C/T	0.30	**5.4E-05**	**6.1E-05**	**0.010**	0.29	0.232	0.239	**0.024**
rs837763	CDT1	C/T	0.44	0.591	0.804	**0.046**	0.56	0.492	0.554	0.533
rs857691	SPTA1	T/C	0.25	0.803	0.492	**5.6E-06**	0.30	0.092	0.226	0.894
rs9604573	GAS6	A/G	0.26	0.482	0.300	0.687	0.24	0.902	0.833	0.425
rs9818758	USP4	A/G	0.17	0.240	0.870	0.780	0.09	0.297	0.297	0.642
rs9914988	ERAL1	G/A	0.21	0.307	0.993	0.058	0.66	0.252	0.080	0.342

*Candidate SNPs selected from NHGRI database based on previous genome-wide significant associations

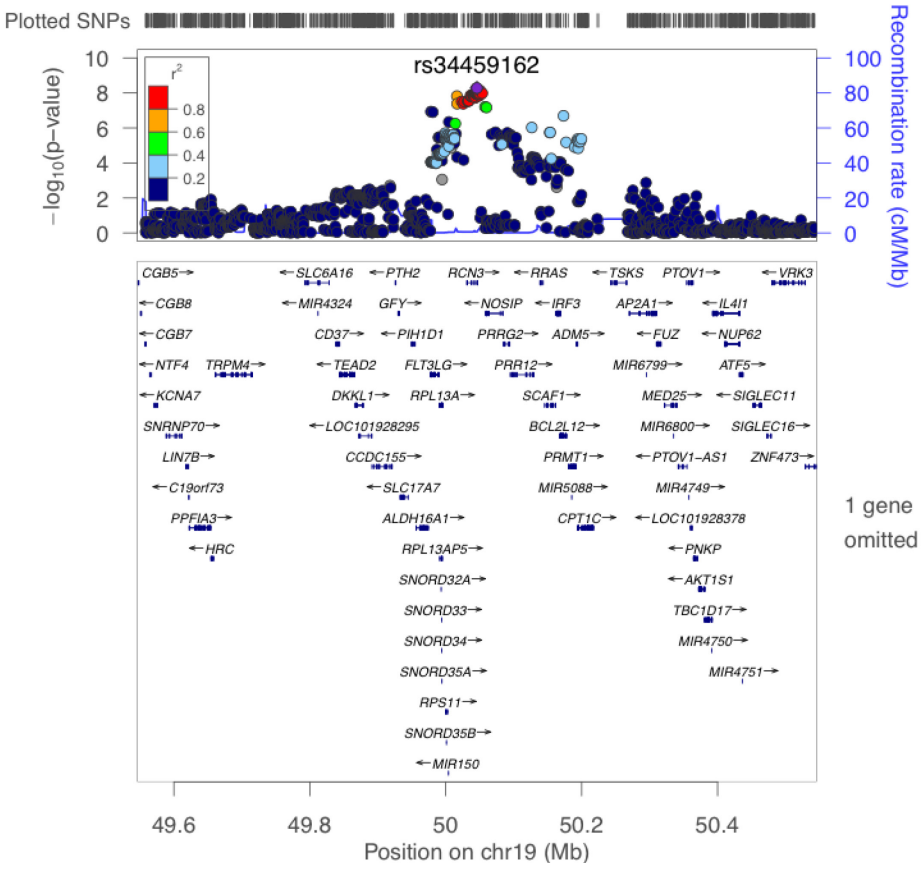
†P-values that reach study-wide significance for fructosamine and glycated albumin, $p<2.7x10^{-4}$ (0.05/(2*2*46)) are in bold; P-values that reach nominal significance ($p<0.05$) for HbA1c are in bold

Figure 3.1. Manhattan plot for GWAS of fructosamine* in whites (N=7,647)



*Fructosamine is log transformed.

Figure 3.2. Regional association plot for rs34459162 and fructosamine* among whites†



*Fructosamine is log transformed.

†Included SNPs with MAF ≥5% and imputation quality (INFO) ≥0.8, insertions and deletions excluded

Supplemental Table 3.1. Characteristics of participants included in the GWAS*

	ARIC		CARDIA	
	Blacks	**Whites**	**Blacks**	**Whites**
N	2,104 (22%)	7,647 (78%)	608 (32%)	1,304 (68%)
Age	56 (6)	57 (6)	45 (4)	46 (3)
Male	782 (37%)	3502 (46%)	230 (38%)	604 (46%)
ARIC site				
Forsyth County, NC	227 (11%)	2300 (30%)	-	-
Jackson, MS	1877 (89%)	0	-	-
Minneapolis, MN	0	2911 (38%)	-	-
Washington County, MD	0	2436 (32%)	-	-
CARDIA site				
Birmingham, AL	-	-	180 (30%)	258 (20%)
Chicago, IL	-	-	128 (21%)	305 (23%)
Minneapolis, MN	-	-	97 (16%)	415 (32%)
Oakland, CA	-	-	203 (33%)	326 (25%)
Fasting glucose (mg/dL)†	109 (25)	104 (17)	95 (11)	95 (11)
HbA1c (%; mmol/mol)‡	5.8 (0.8); 40 (8.7)	5.4 (0.5); 36 (5.5)	5.5 (0.5); 37 (5.5)	5.3 (0.3); 34 (3.3)
Fructosamine (μmol/L)	238 (35)	227 (23)	227 (21)	225 (18)
Percent glycated albumin (%)	14 (3)	13 (2)	13 (1)	13 (1)
Serum albumin (g/dL)	4.1 (0.3)	4.2 (0.3)	4.1 (0.3)	4.3 (0.3)
Undiagnosed diabetes§	226 (11%)	418 (5%)	-	-

*Continuous variables shown as mean (SD) and categorical variables shown as n (%)

†ARIC fasting glucose N=9,574

‡ARIC HbA1c N=9,626

§Undiagnosed diabetes defined as fasting glucose ≥126 mg/dL if fasting for ≥8 hours or fasting glucose >200 if not fasting for ≥8 hours

Supplemental Table 3.2. Power for replication of ARIC genome-wide significant results in CARDIA and across race in ARIC.*

SNP	Closest Gene	Outcome	Cohort	Race	N	Mean (SD)	RAF†	Beta	Power
rs34459162	RCN3	Fructosamine (umol/L)	CARDIA	White	1304	5.41 (0.08)	0.09	0.02	0.96
rs1260326	GCKR	Percent glycated albumin (%)	CARDIA	White	1304	2.53 (0.09)	0.43	0.01	0.80
			ARIC	Black	2104	2.60 (0.14)	0.14	0.000	0.05
rs2438321	CNTN5	Fructosamine (umol/L)	CARDIA	Black	608	5.42 (0.09)	0.06	0.006	0.09
			ARIC	White	7647	5.41 (0.09)	0.24	0.001	0.09

*Power calculations were done in Quanto (http://biostats.usc.edu/Quanto.html). Quanto assumptions: continuous trait, gene only, additive inheritance, alpha=0.05 for replication.
†RAF=risk allele frequency

Supplemental Table 3.3. Association result for top fructosamine and percent glycated albumin SNPs after excluding samples with undiagnosed diabetes in ARIC (N whites=7,229; N blacks=1,878) *,†,‡

SNP	Closest Gene	A1/ A2	Race	A1 freq§	Fructosamine (µmol/L)		Percent glycated albumin (%)	
					Beta (SE)	P-value	Beta (SE)	P-value
rs34459162	RCN3	C/T	White	0.08	-0.02 (0.003)	**6.8E-10**	-0.02 (0.003)	**1.3E-8**
			Black	-	-	-	-	-
rs1260326	GCKR	T/C	White	0.41	-0.002 (0.001)	0.10	-0.01 (0.002)	**4.7E-9**
			Black	0.15	-0.003 (0.004)	0.52	-0.005 (0.005)	0.32
rs2438321	CNTN5	G/A	White	0.24	-0.001 (0.002)	0.74	0.002 (0.002)	0.30
			Black	0.11	0.02 (0.005)	4.2E-5	0.01 (0.005)	0.18
rs59443763	PRKCA	C/T	White	-	-	-	-	-
			Black	0.06	0.02 (0.006)	0.01	0.03 (0.007)	1.1E-4

*Undiagnosed diabetes defined as fasting glucose ≥126 mg/dL if fasting for ≥8 hours or fasting glucose >200 if not fasting for ≥8 hours

†Genome-wide significant results are in bold

‡Fructosamine and percent glycated albumin are log transformed

§A1 is minor allele in whites

Supplemental Table 3.4. Association results for top SNPs with serum albumin (N=7,586) *,†

SNP	Closest Gene	A1/A2‡	Race	A1 freq	Beta (SE)	P-value
rs34459162	*RCN3*	C/T	White	0.08	-0.003 (0.002)	0.17
			Black	-	-	-
rs1260326	*GCKR*	C/T	White	0.40	0.003 (0.001)	0.01
			Black	0.14	0.002 (0.003)	0.57
rs2438321	*CNTN5*	G/A	White	0.24	-0.002 (0.001)	0.15
			Black	0.11	0.01 (0.004)	**0.002**
rs59443763	*PRKCA*	C/T	White	-	-	-
			Black	0.06	-0.004 (0.005)	0.41

*Bonferroni corrected significance threshold (0.05/(4 SNPs * 2 races) = 0.006) results are in bold

†Serum albumin is log transformed

‡A1 is the minor allele in whites

Supplemental Table 3.5. Genetic correlations between biomarkers among whites in ARIC

	Fructosamine (umol/L)	**Percent glycated albumin (%)**	**Total glycated albumin (g/dL)**	**Fasting glucose (mg/dL)**	**HbA1c (%)**
Fructosamine (umol/L)	--	1.17 (0.21) **1.7E-08**	0.92 (0.09) **1.1E-26**	0.81 (0.43) 0.06	0.47 (0.32) 0.14
Percent glycated albumin (%)		--	1.06 (0.08) **1.8E-40**	0.58 (0.51) 0.26	0.21 (0.47) 0.66
Total glycated albumin (g/dL)			--	0.68 (0.44) 0.12	0.36 (0.33) 0.27
Fasting glucose (mg/dL)				--	0.47 (0.42) 0.26
HbA1c (%)					--

Supplemental Table 3.6. Beta and standard errors for associations between fasting glucose known genetic determinants and fructosamine and percent glycated albumin in ARIC*†

		Whites				Blacks			
			Fructo samine	**Percent Glycated Albumin**	**Fasting Glucose**		**Fructos amine**	**Percent Glycated Albumin**	**Fasting Glucose**
SNP	A1/ A2	A1 freq	Beta (SE)	Beta (SE)	Beta (SE)	A1 freq	Beta (SE)	Beta (SE)	Beta (SE)
rs10747083	G/A	0.33	0 (0.002)	0.001 (0.003)	-0.455 (0.311)	0.16	-0.008 (0.006)	-0.011 (0.007)	-0.053 (1.278)
rs10830963	G/C	0.28	0.006 (0.002)	0.008 (0.003)	**1.577 (0.309)**	0.07	0.011 (0.007)	0.018 (0.008)	1.33 (1.529)
rs10885122	T/G	0.12	0.001 (0.002)	-0.001 (0.004)	0.13 (0.418)	0.68	0.005 (0.004)	0.002 (0.005)	1.069 (0.858)
rs11071657	G/A	0.38	-0.001 (0.002)	-0.001 (0.003)	0.184 (0.287)	0.12	0 (0.006)	0 (0.007)	0.988 (1.273)
rs11558471	G/A	0.32	-0.005 (0.002)	**-0.011 (0.003)**	**-1.256 (0.293)**	0.09	-0.007 (0.007)	-0.01 (0.008)	0.641 (1.412)
rs11603334	A/G	0.16	-0.006 (0.002)	-0.006 (0.003)	-0.266 (0.37)	0.06	-0.002 (0.008)	0.001 (0.009)	0.513 (1.678)
rs11605924	A/C	0.47	-0.001 (0.001)	-0.001 (0.002)	-0.118 (0.27)	0.84	0.003 (0.006)	0.002 (0.007)	0.885 (1.194)
rs11708067	G/A	0.23	-0.003 (0.002)	-0.01 (0.003)	**-0.836 (0.324)**	0.16	-0.016 (0.005)	-0.018 (0.006)	**-2.374 (1.071)**
rs11715915	T/C	0.3	-0.002 (0.002)	0.001 (0.003)	**-0.596 (0.293)**	0.23	0.008 (0.004)	0.005 (0.005)	0.968 (0.949)
rs11920090	A/T	0.13	-0.006 (0.002)	-0.011 (0.004)	-0.375 (0.396)	0.35	-0.003 (0.004)	-0.005 (0.005)	-0.891 (0.819)
rs13179048	A/C	0.31	0.003 (0.002)	0.004 (0.003)	-0.197 (0.298)	0.08	0.009 (0.007)	0.005 (0.008)	0.397 (1.444)
rs1371614	T/C	0.25	-0.001 (0.002)	-0.001 (0.003)	0.131 (0.326)	0.36	0.005 (0.004)	0.007 (0.005)	1.056 (0.844)
rs143399767	C/A	0.01	-0.003 (0.007)	-0.014 (0.012)	-0.725 (1.246)	0.003	NA	NA	NA
rs1483121	A/G	0.13	0.002 (0.003)	0.004 (0.004)	-0.148 (0.475)	0.04	0.009 (0.012)	0.011 (0.014)	0.146 (2.688)
rs16913693	G/T	0.03	-0.002 (0.004)	-0.002 (0.007)	-1.272 (0.796)	0.25	0.001 (0.004)	0.002 (0.005)	-0.106 (0.905)
rs174550	C/T	0.33	0.002 (0.002)	-0.003 (0.003)	-0.178 (0.287)	0.08	-0.006 (0.007)	-0.012 (0.008)	-0.128 (1.501)
rs17762454	T/C	0.26	0.003 (0.002)	0.003 (0.003)	**1.014 (0.315)**	0.16	0.009 (0.007)	0.002 (0.008)	-0.374 (1.38)
rs2191349	G/T	0.46	-0.002 (0.001)	-0.002 (0.003)	**-0.646 (0.273)**	0.44	-0.002 (0.004)	-0.002 (0.004)	**-2.274 (0.786)**
rs2293941	A/G	0.22	0.004 (0.002)	0.009 (0.003)	**0.73 (0.332)**	0.17	-0.004 (0.005)	-0.003 (0.006)	-0.304 (1.054)
rs2302593	G/C	0.5	0 (0.002)	0.002 (0.003)	**-0.684 (0.305)**	0.3	-0.002 (0.004)	0 (0.005)	-0.03 (0.926)
rs2657879	G/A	0.18	0 (0.002)	0 (0.003)	0.357 (0.351)	0.06	-0.015 (0.008)	-0.011 (0.009)	**-3.543 (1.695)**
rs2722425	T/C	0.12	0.006 (0.002)	0.01 (0.004)	0.116 (0.422)	0.37	-0.004 (0.004)	0.004 (0.005)	0.689 (0.832)
rs340874	T/C	0.45	-0.001 (0.002)	-0.001 (0.003)	0.231 (0.276)	0.17	0.005 (0.005)	-0.002 (0.006)	1.506 (1.056)

rs35767	G/A	0.16	0.001 (0.002)	0.002 (0.004)	-0.192 (0.39)	0.55	0.003 (0.004)	0.007 (0.004)	**1.918 (0.805)**
rs3736594	C/A	0.26	0.003 (0.002)	0.005 (0.003)	0.356 (0.308)	0.43	0.007 (0.004)	0.01 (0.004)	-0.4 (0.803)
rs3783347	T/G	0.22	0.001 (0.002)	0.002 (0.003)	0.189 (0.333)	0.06	0.006 (0.008)	0.005 (0.009)	0.121 (1.608)
rs3829109	A/G	0.29	-0.005 (0.002)	-0.004 (0.003)	-0.458 (0.325)	0.18	0.003 (0.005)	0.005 (0.006)	0.783 (1.084)
rs4506565	T/A	0.31	0.006 (0.002)	**0.011 (0.003)**	**1.232 (0.293)**	0.44	0.006 (0.004)	0.004 (0.004)	-0.169 (0.79)
rs4607517	A/G	0.17	**0.007 (0.002)**	0.01 (0.003)	**1.275 (0.366)**	0.11	0.011 (0.006)	0.013 (0.008)	2.416 (1.374)
rs4841132	A/G	0.09	-0.003 (0.003)	-0.004 (0.004)	**0.947 (0.481)**	0.13	0.002 (0.007)	0.008 (0.008)	**2.718 (1.393)**
rs560887	T/C	0.3	-0.003 (0.002)	-0.008 (0.003)	**-1.178 (0.297)**	0.05	0.002 (0.008)	-0.005 (0.01)	-0.203 (1.777)
rs576674	G/A	0.16	0.002 (0.002)	0.002 (0.003)	0.364 (0.373)	0.61	0.005 (0.004)	0.005 (0.005)	1.308 (0.823)
rs6048205	G/A	0.05	0.001 (0.003)	0 (0.006)	-1.134 (0.626)	0.18	-0.011 (0.005)	-0.007 (0.006)	**-2.05 (1.032)**
rs6072275	A/G	0.16	0.001 (0.002)	0.004 (0.003)	**0.902 (0.371)**	0.08	0.01 (0.007)	0.013 (0.008)	2.603 (1.463)
rs6943153	T/C	0.31	-0.001 (0.002)	-0.002 (0.003)	0.419 (0.296)	0.7	0.007 (0.004)	0.001 (0.005)	0.424 (0.837)
rs7034200	A/C	0.49	0.002 (0.002)	0.005 (0.003)	**0.584 (0.274)**	0.61	0.005 (0.004)	0.011 (0.005)	1.354 (0.838)
rs7651090	G/A	0.31	0.004 (0.002)	0.006 (0.003)	**0.88 (0.295)**	0.56	0.004 (0.004)	0.002 (0.004)	0.693 (0.816)
rs7708285	G/A	0.3	0.002 (0.002)	-0.001 (0.003)	0.119 (0.33)	0.15	0.004 (0.007)	0.005 (0.008)	-1.902 (1.359)
rs780094	T/C	0.4	-0.003 (0.002)	**-0.01 (0.003)**	**-1.075 (0.277)**	0.18	0.002 (0.005)	0 (0.006)	-0.135 (1.035)
rs7944584	T/A	0.27	0.001 (0.002)	0.001 (0.003)	**-0.603 (0.306)**	0.04	0 (0.009)	0.009 (0.01)	-0.539 (1.865)
rs9368222	A/C	0.27	0.002 (0.002)	0.005 (0.003)	0.336 (0.307)	0.2	0 (0.005)	-0.001 (0.006)	0.412 (1.011)

*Candidate SNPs selected from NHGRI database based on previous genome-wide significant associations
†P-values that reach study-wide significance for fructosamine and glycated albumin, $p<3.0x10^{-4}$ (0.05/(2*2*41)) are in bold; P values that reach nominal significance ($p<0.05$) for fasting glucose in bold

Supplemental Table 3.7. Beta and standard errors for associations between HbA1c known genetic determinants and fructosamine and percent glycated albumin in ARIC*

		Whites				Blacks			
			Fructosamine	**Percent Glycated Albumin**	**HbA1c**		**Fructosamine**	**Percent Glycated Albumin**	**HbA1c**
SNP	A1/A2	A1 freq	Beta (SE)	Beta (SE)	Beta (SE)	A1 freq	Beta (SE)	Beta (SE)	Beta (SE)
rs1046896	T/C	0.31	-0.003 (0.002)	-0.005 (0.003)	0.017 (0.009)	0.24	0.002 (0.004)	0.004 (0.005)	0.017 (0.031)
rs10774625	G/A	0.48	-0.001 (0.002)	0.006 (0.003)	**0.024 (0.009)**	0.09	-0.008 (0.006)	-0.011 (0.008)	-0.03 (0.045)
rs10823343	G/A	0.26	-0.001 (0.002)	-0.005 (0.003)	**-0.047 (0.01)**	0.45	-0.001 (0.004)	0.003 (0.005)	-0.003 (0.027)
rs10830963	G/C	0.28	0.006 (0.002)	0.008 (0.003)	**0.026 (0.009)**	0.07	0.011 (0.007)	0.018 (0.008)	-0.061 (0.05)
rs11248914	C/T	0.35	-0.004 (0.002)	-0.003 (0.003)	-0.017 (0.009)	0.36	0.007 (0.004)	0.004 (0.004)	0.019 (0.027)
rs11558471	G/A	0.32	-0.005 (0.002)	**-0.011 (0.003)**	**-0.023 (0.009)**	0.09	-0.007 (0.007)	-0.01 (0.008)	0.001 (0.046)
rs11603334	A/G	0.16	-0.006 (0.002)	-0.006 (0.003)	-0.014 (0.011)	0.06	-0.002 (0.008)	0.001 (0.009)	0.019 (0.055)
rs11708067	G/A	0.23	-0.003 (0.002)	-0.01 (0.003)	**-0.021 (0.01)**	0.16	-0.016 (0.005)	-0.018 (0.006)	-0.063 (0.035)
rs11954649	G/C	0.00	NA	NA	NA	0.05	-0.002 (0.009)	0.013 (0.01)	0.082 (0.062)
rs11964178	G/A	0.43	0 (0.002)	0 (0.003)	-0.003 (0.008)	0.36	-0.005 (0.004)	-0.001 (0.005)	0.02 (0.027)
rs12621844	C/T	0.39	0.001 (0.002)	0.003 (0.003)	-0.008 (0.009)	0.84	-0.001 (0.005)	-0.001 (0.006)	-0.015 (0.035)
rs12819124	A/C	0.47	0 (0.001)	0 (0.002)	**-0.028 (0.008)**	0.21	0.003 (0.005)	-0.001 (0.005)	-0.046 (0.031)
rs13134327	A/G	0.32	0 (0.002)	-0.001 (0.003)	**0.019 (0.009)**	0.3	-0.005 (0.004)	0.001 (0.005)	-0.008 (0.028)
rs1402837	T/C	0.22	0.004 (0.002)	**0.012 (0.003)**	**0.052 (0.01)**	0.31	-0.003 (0.004)	-0.002 (0.005)	-0.013 (0.029)
rs1558902	A/T	0.41	-0.001 (0.002)	-0.006 (0.003)	0.013 (0.008)	0.11	-0.011 (0.006)	-0.009 (0.007)	-0.027 (0.041)
rs17509001	C/T	0.14	-0.002 (0.002)	-0.005 (0.004)	**0.026 (0.012)**	0.09	0.003 (0.006)	-0.005 (0.007)	-0.011 (0.044)
rs17533903	A/G	0.21	0.001 (0.002)	0.004 (0.003)	**0.036 (0.01)**	0.24	-0.004 (0.004)	-0.006 (0.005)	-0.012 (0.03)
rs17747324	C/T	0.23	**0.007 (0.002)**	**0.013 (0.003)**	**0.037 (0.01)**	0.07	0.006 (0.008)	0.003 (0.009)	0.007 (0.054)
rs1800562	A/G	0.06	-0.003 (0.003)	-0.008 (0.005)	**-0.058 (0.017)**	0.01	0.003 (0.016)	-0.008 (0.019)	-0.165 (0.114)
rs198846	A/G	0.16	0.002 (0.002)	0.004 (0.003)	-0.018 (0.011)	0.88	-0.004 (0.006)	-0.007 (0.007)	-0.045 (0.04)
rs2110073	T/C	0.1	-0.002 (0.002)	0 (0.004)	**0.043 (0.014)**	0.42	-0.005 (0.004)	-0.004 (0.004)	-0.037 (0.027)
rs2383208	G/A	0.18	-0.005 (0.002)	-0.008 (0.003)	**-0.029 (0.011)**	0.19	0.005 (0.005)	0.002 (0.005)	0.017 (0.033)
rs2408955	G/T	0.48	-0.001 (0.001)	0 (0.003)	**-0.026 (0.008)**	0.63	0.001 (0.004)	0.001 (0.005)	-0.013 (0.027)

rs267738	G/T	0.21	-0.001 (0.002)	-0.002 (0.003)	-0.007 (0.01)	0.04	-0.002 (0.009)	-0.003 (0.011)	-0.066 (0.064)
rs2779116	T/C	0.27	0.001 (0.002)	-0.001 (0.003)	**0.041 (0.009)**	0.22	-0.007 (0.004)	-0.007 (0.005)	-0.002 (0.031)
rs282587	G/A	0.12	0.001 (0.002)	0.004 (0.004)	**0.049 (0.013)**	0.69	-0.008 (0.004)	-0.007 (0.005)	-0.033 (0.029)
rs3824065	T/C	0.42	-0.006 (0.002)	-0.008 (0.003)	-0.014 (0.009)	0.24	0.009 (0.005)	0.009 (0.006)	-0.003 (0.036)
rs4607517	A/G	0.17	**0.007 (0.002)**	0.01 (0.003)	**0.045 (0.011)**	0.11	0.011 (0.006)	0.013 (0.008)	0.052 (0.045)
rs4737009	A/G	0.24	-0.003 (0.002)	-0.002 (0.003)	**0.022 (0.01)**	0.46	-0.004 (0.004)	-0.003 (0.004)	-0.024 (0.026)
rs4745982	G/T	0.07	0.004 (0.003)	0.006 (0.005)	**-0.08 (0.018)**	0.1	-0.001 (0.007)	0.001 (0.009)	-0.026 (0.052)
rs4820268	G/A	0.46	-0.002 (0.001)	-0.002 (0.003)	**0.021 (0.008)**	0.72	0.008 (0.004)	0.011 (0.005)	0.04 (0.03)
rs560887	T/C	0.3	-0.003 (0.002)	-0.008 (0.003)	**-0.027 (0.009)**	0.95	0.002 (0.008)	-0.005 (0.01)	0.01 (0.057)
rs579459	C/T	0.23	0.002 (0.002)	0.003 (0.003)	**0.033 (0.01)**	0.13	-0.007 (0.005)	-0.004 (0.006)	0.02 (0.038)
rs592423	A/C	0.46	0.001 (0.001)	0.002 (0.002)	0.015 (0.008)	0.61	-0.004 (0.004)	-0.002 (0.004)	0.025 (0.026)
rs6474359	C/T	0.03	0 (0.005)	0.005 (0.008)	**-0.089 (0.025)**	0.27	0.008 (0.004)	0.01 (0.005)	-0.005 (0.03)
rs6980507	A/G	0.39	-0.002 (0.002)	0 (0.003)	**0.02 (0.008)**	0.48	-0.002 (0.004)	-0.003 (0.004)	-0.011 (0.025)
rs7040409	G/C	0.07	-0.002 (0.003)	-0.002 (0.005)	**-0.057 (0.017)**	0.26	-0.009 (0.004)	-0.014 (0.005)	-0.047 (0.03)
rs7616006	G/A	0.43	0.001 (0.002)	-0.002 (0.003)	-0.011 (0.009)	0.36	0 (0.004)	0 (0.005)	-0.006 (0.028)
rs761772	C/T	0.13	0.004 (0.002)	0.007 (0.004)	**0.036 (0.014)**	0.13	-0.004 (0.006)	-0.006 (0.007)	0.007 (0.04)
rs7756992	G/A	0.27	0.002 (0.002)	0.005 (0.003)	0.01 (0.009)	0.42	0.001 (0.004)	0 (0.004)	0 (0.026)
rs8192675	C/T	0.3	**-0.007 (0.002)**	**-0.011 (0.003)**	**-0.023 (0.009)**	0.29	-0.005 (0.004)	-0.006 (0.005)	**-0.063 (0.028)**
rs837763	C/T	0.44	0.001 (0.002)	0.001 (0.003)	**-0.018 (0.009)**	0.56	0.003 (0.004)	0.003 (0.005)	-0.017 (0.028)
rs857691	T/C	0.25	0 (0.002)	-0.002 (0.003)	**0.044 (0.01)**	0.3	-0.007 (0.004)	-0.006 (0.005)	-0.004 (0.029)
rs9604573	A/G	0.26	-0.001 (0.002)	-0.003 (0.003)	0.004 (0.01)	0.24	0 (0.005)	0.001 (0.006)	0.026 (0.033)
rs9818758	A/G	0.17	-0.002 (0.002)	0.001 (0.003)	-0.003 (0.011)	0.09	0.007 (0.006)	0.008 (0.008)	0.021 (0.046)
rs9914988	G/A	0.21	0.002 (0.002)	0 (0.003)	-0.019 (0.01)	0.66	0.005 (0.004)	0.008 (0.005)	0.026 (0.028)

*Candidate SNPs selected from NHGRI database based on previous genome-wide significant associations
†P-values that reach study-wide significance for fructosamine and glycated albumin, $p<2.7x10^{-4}$ (0.05/(2*2*46)) are in bold; P-values that reach nominal significance ($p<0.05$) for HbA1c are in bold

Supplemental Table 3.8. Pearson's correlation coefficients among glycemic biomarkers.†

	Fasting glucose (mg/dL)			HbA1c (%)			Fructosamine (µmol/L)			Percent glycated albumin (%)		
	W	B	Total	W	B	Total	W	B	Total	W	B	Total
HbA1c (%; mmol/mol)	0.69	0.77	0.72									
Fructosamine (µmol/L)	0.52	0.68	0.58	0.45	0.61	0.53						
Percent glycated albumin (%)	0.54	0.61	0.59	0.53	0.69	0.61	0.76	0.84	0.79			
Total glycated albumin (g/dL)	0.53	0.63	0.59	0.47	0.63	0.55	0.85	0.91	0.87	0.93	0.94	0.93

†N=1990 Blacks, N=7465 Whites; B=Black, W=White

Supplemental Table 3.9. Percent variance of HbA1c explained by known glycemic and nonglycemic SNPs associated with HbA1c*

Source	Gene	SNP	MAF	Beta	Var(SNP)	Variance explained†
Erythrocytic						
Wheeler	TMEM79	rs12132919	0.29	0.01	0.41	4.2E-05
Wheeler	SPTA1	rs857691	0.25	0.04	0.38	0.003
Wheeler	HK1	rs4745982	0.07	-0.08	0.13	0.003
Wheeler	CNTN5	rs11224302	0.10	-0.05	0.18	0.002
Wheeler	ATXN2	rs10774625	0.48	0.02	0.50	0.001
Wheeler	SENP1	rs2408955	0.48	0.03	0.50	0.001
Wheeler	ITFG3	rs11248914	0.35	-0.02	0.45	0.001
Wheeler	CDH3	rs4783565	0.27	0.01	0.39	1.6E-04
Wheeler	CDT1	rs837763	0.44	0.02	0.49	0.001
Wheeler	ERAL1	rs9914988	0.21	0.02	0.33	4.9E-04
Wheeler	MYO9B	rs17533903	0.21	0.04	0.33	0.002
Wheeler	TMPRSS6	rs4820268	0.46	-0.02	0.50	0.001
Wheeler	SYN2	rs7616006	0.43	-0.01	0.49	2.5E-04
Wheeler	C6orf183	rs11964178	0.43	0.00	0.49	2.0E-05
Wheeler	CITED2	rs592423	0.45	-0.02	0.50	4.6E-04
Wheeler + Soranzo	HFE	rs1800562	0.06	-0.06	0.12	0.002
Wheeler	HFE	rs198846	0.16	0.02	0.26	3.3E-04
Wheeler + Soranzo + Chen	ANK1	rs4737009	0.24	0.02	0.37	0.001
Wheeler	SLC20A2	rs6980507	0.39	0.02	0.48	0.001
Wheeler	C9orf47	rs7040409	0.07	-0.06	0.12	0.002
Soranzo + Chen‡		rs1046896	0.31	0.02	0.43	0.001
Chen	MYO9B	rs11667918	0.28	-0.02	0.40	0.001
Soranzo		rs16926246	0.12	-0.09	0.21	0.006
Soranzo		rs2779116	0.27	0.04	0.39	0.003
Chen	TMEM79	rs6684514	0.29	0.01	0.41	7.7E-05
Soranzo		rs7998202	0.12	0.05	0.22	0.003
Soranzo		rs855791	0.43	0.02	0.49	0.001
Chen	HBS1L/MYB	rs9399137	0.27	-0.01	0.39	2.2E-04
Chen	CYBA	rs9933309	0.28	-0.01	0.41	9.5E-05
Total						**0.034**
Glycemic						
Wheeler	TCF7L2	rs17747324	0.23	0.04	0.36	0.002
Wheeler	KCNQ1	rs2237896	0.05	-0.05	0.09	0.001
Wheeler	FADS2	rs174577	0.34	0.00	0.45	4.4E-06
Wheeler	ARAP1	rs11603334	0.16	-0.01	0.27	2.0E-04
Wheeler	MTNR1B	rs10830963	0.28	0.03	0.40	0.001
Wheeler	PDX1	rs11619319	0.22	0.03	0.34	0.001
Wheeler	KL	rs576674	0.16	0.00	0.27	8.5E-06
Wheeler	G6PC2	rs13387347	0.46	0.02	0.50	0.001
Wheeler	G6PC2	rs560887	0.30	0.03	0.42	0.001
Wheeler	ADCY5	rs11708067	0.23	-0.02	0.35	0.001
Wheeler	SLC2A2	rs8192675	0.30	-0.02	0.42	0.001
Wheeler	FREM3	rs13134327	0.32	0.02	0.44	0.001
Wheeler	CDKAL1	rs7756992	0.27	0.01	0.40	1.5E-04
Wheeler	DGKB	rs2191349	0.47	0.01	0.50	3.6E-04
Wheeler	GCK	rs4607517	0.17	0.04	0.28	0.002
Wheeler	GCK	rs3824065	0.42	-0.01	0.49	3.6E-04
Wheeler	SLC30A8	rs11558471	0.32	-0.02	0.43	0.001
Wheeler	ABO	rs579459	0.23	0.03	0.35	0.002

Wheeler	MTAP	rs2383208	0.18	-0.03	0.30	0.001
Soranzo	MTNR1B	rs1387153	0.29	0.03	0.41	0.001
Soranzo + Chen	GCK	rs1799884	0.17	0.04	0.28	0.002
Chen	G6PC2/ABCB11	rs3755157	0.11	-0.01	0.20	3.6E-05
Soranzo	G6PC2	rs552976	0.35	-0.03	0.45	0.002
Chen	CDKAL1	rs7772603	0.27	0.01	0.39	3.0E-04
Total						**0.021**
Unclassified						
Wheeler	CERS2	rs267738	0.21	-0.01	0.33	5.8E-05
Wheeler	SYF2	rs2375278	0.17	-0.01	0.28	2.5E-04
Wheeler	HK1	rs10823343	0.26	-0.05	0.38	0.003
Wheeler	PHB2	rs2110073	0.10	0.04	0.18	0.001
Wheeler	ATP11A	rs282587	0.12	-0.05	0.21	0.002
Wheeler	GAS6	rs9604573	0.26	0.00	0.38	2.5E-05
Wheeler	FTO	rs1558902	0.41	0.01	0.48	3.1E-04
Wheeler	TMC6	rs2073285	0.21	0.01	0.33	9.8E-05
Wheeler	FN3KRP	rs1046896	0.31	0.02	0.43	0.001
Wheeler	MYO9B	rs11086054	0.31	-0.02	0.42	4.0E-04
Wheeler	SCRN3	rs17256082	0.34	0.02	0.45	0.001
Wheeler	ATAD2B	rs17509001	0.14	0.03	0.24	0.001
Wheeler	FOXN2	rs12621844	0.39	0.01	0.47	1.2E-04
Wheeler	FNDC3B	rs4894799	0.38	0.00	0.47	3.6E-05
Wheeler	USP4	rs9818758	0.17	0.00	0.28	1.1E-05
Wheeler + Soranzo§	ANK1	rs6474359	0.03	-0.09	0.06	0.002
Wheeler	KLF4	rs1467311	0.34	0.00	0.45	2.8E-07
Total						0.012

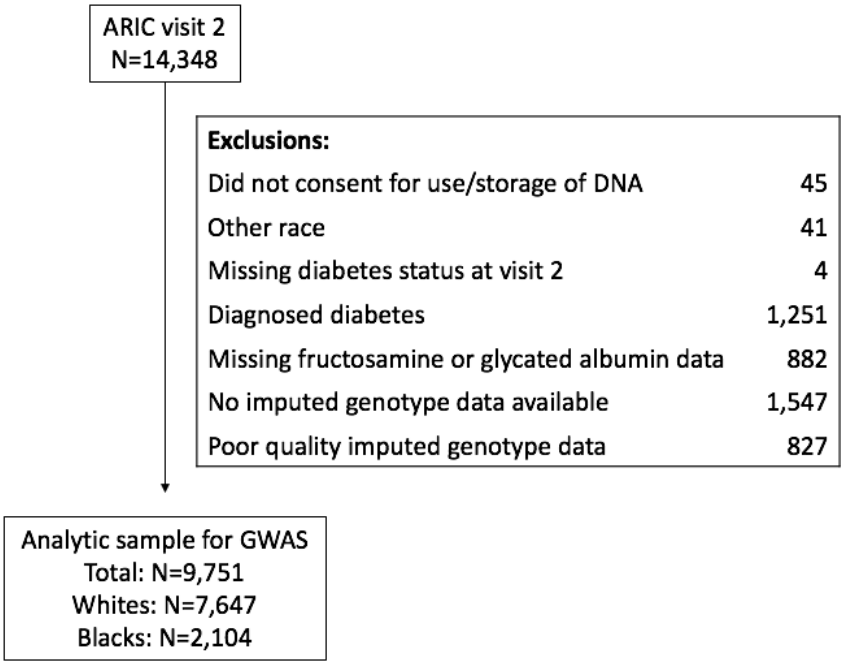
*All variants come from Chen 2014[12] and Soranzo 2010[11] which identified "glycemic" and "nonglycemic" SNPs. Wheeler 2017 designated SNPs as "erythrocytic", "glycemic" or "unclassified"

†Variance explained was calculated using the equation: $R_i^2 = b_i^2$ x var(SNP_i)/var(y) where b_i = the effect size of the association between the SNP_i and the phenotype y, var(SNP_i) is 2 x MAF_{SNPi} x (1-MAF_{SNPi}) and var(y) is the variance in the phenotype (variance in HbA1c in ARIC=0.25)

‡Chen SNP: rs1046875, r^2=1 with rs1046896

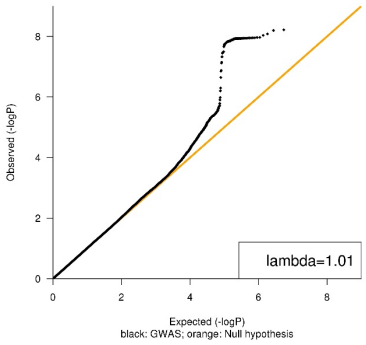
§"unclassified" per Wheeler, "nonglycemic" per Soranzo

Supplemental Figure 3.1. Selection of participants

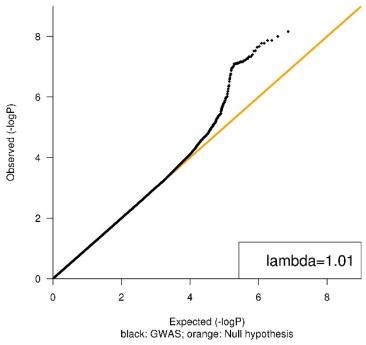


Diagnosed diabetes: self-reported physician diagnosis or taking diabetes medication

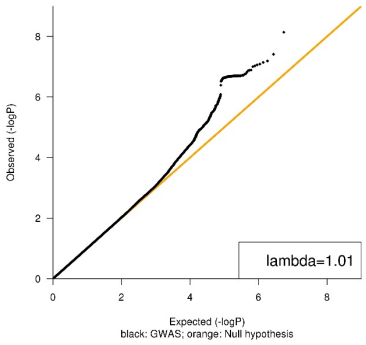
Supplemental Figure 3.2. QQ plot of fructosamine in whites.



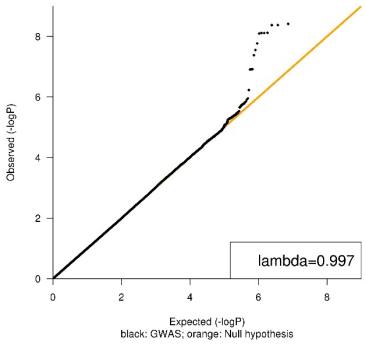
Supplemental Figure 3.3. QQ plot of fructosamine in blacks.



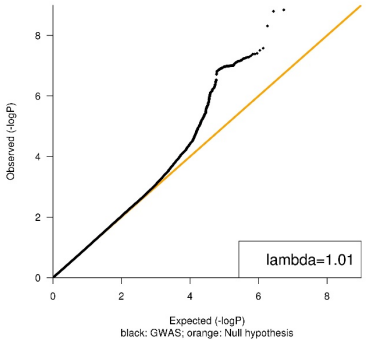
Supplemental Figure 3.4. QQ plot of percent glycated albumin whites.



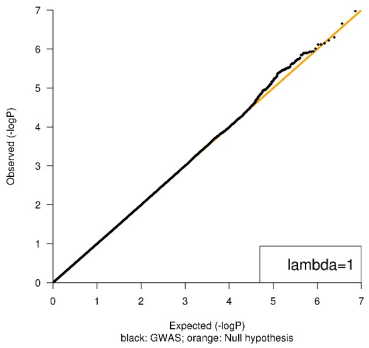
Supplemental Figure 3.5. QQ plot of percent glycated albumin blacks.



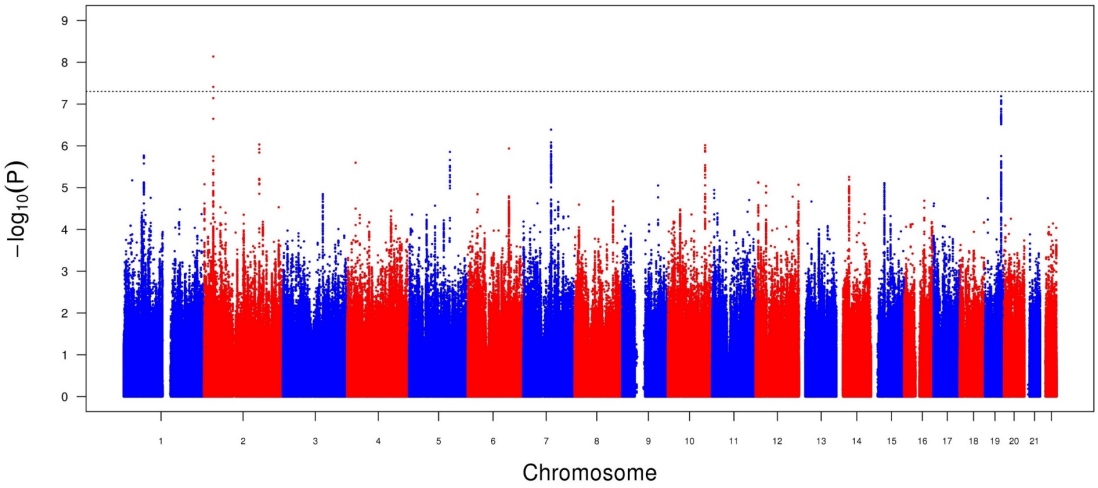
Supplemental figure 3.6. QQ plot of total glycated albumin in whites.



Supplemental figure 3.7. QQ plot of total glycated albumin in blacks.

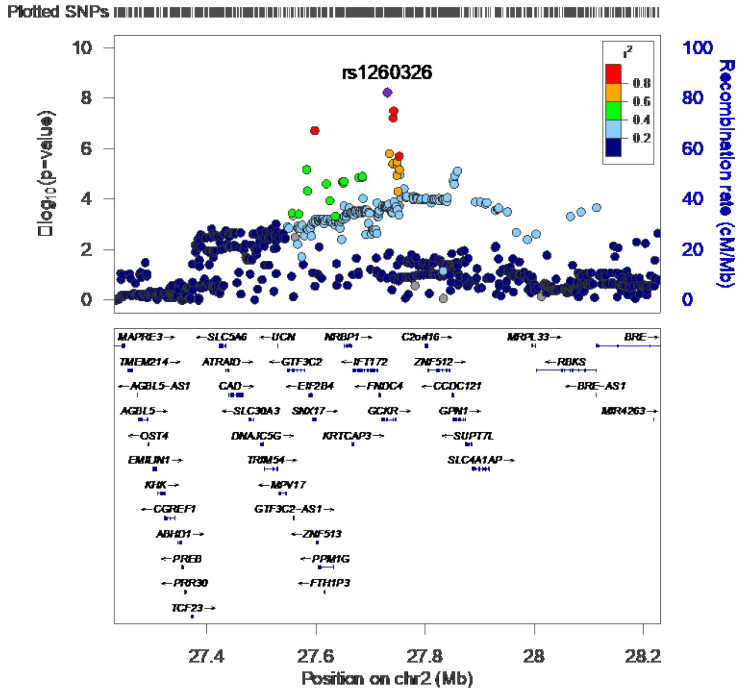


Supplemental Figure 3.8. Manhattan plot for GWAS of percent glycated albumin* in whites (N=7,647).

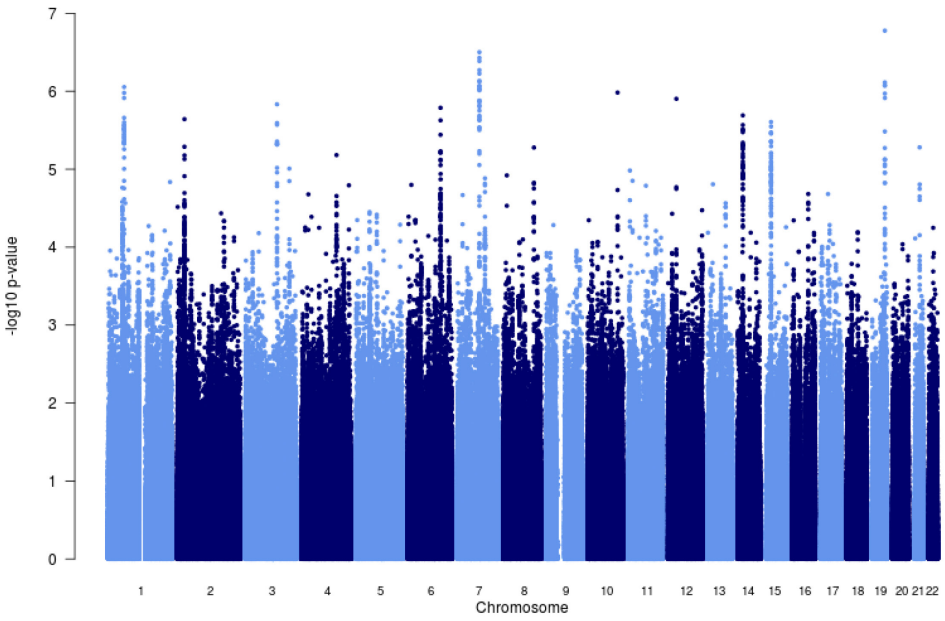


*Percent glycated albumin is natural log transformed

Supplemental Figure 3.9. Regional association plot for rs1260326 and percent glycated albumin among whites.

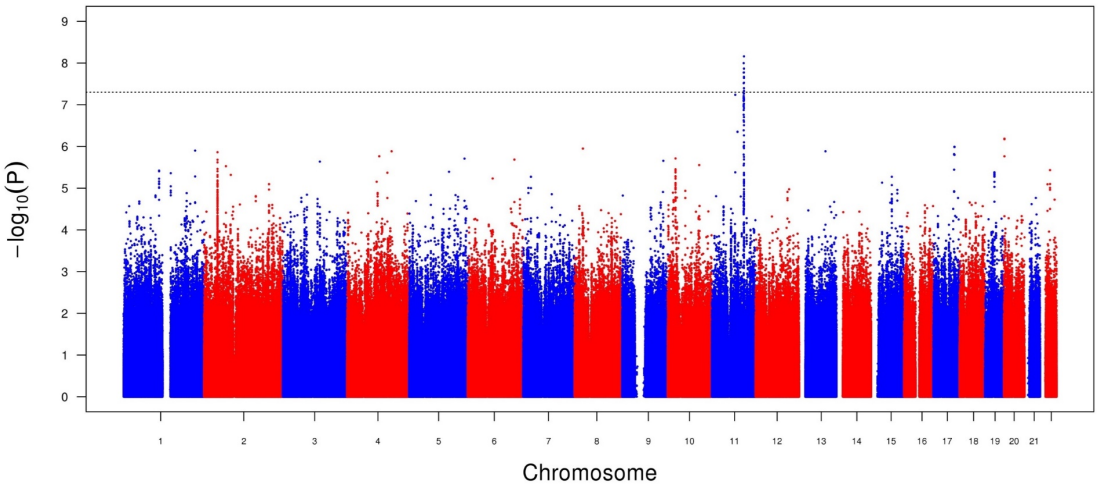


Supplemental Figure 3.10. Manhattan plot for meta-analysis of black and white ARIC participants for percent glycated albumin (N=9,751)



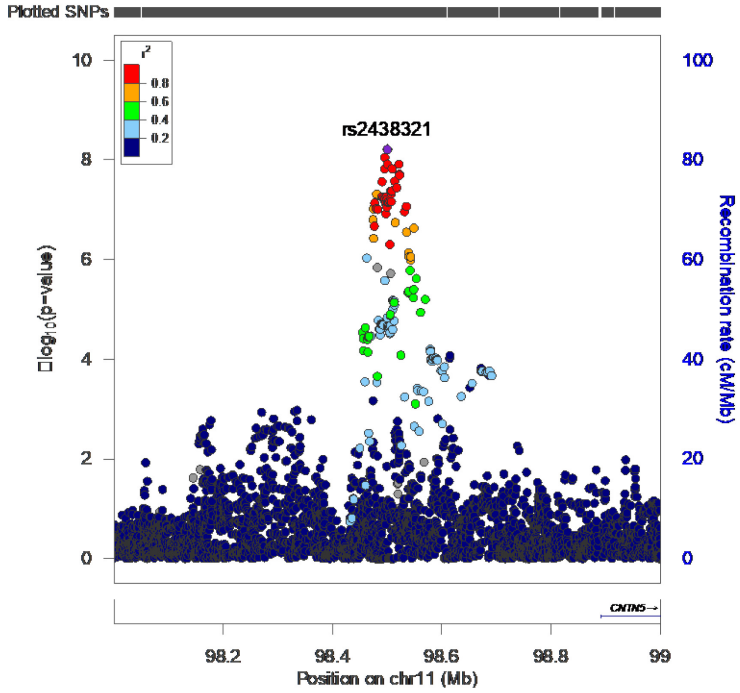
*Percent glycated albumin is natural log transformed.

Supplemental Figure 3.11. Manhattan plot for GWAS of fructosamine* in blacks (N=2,104)

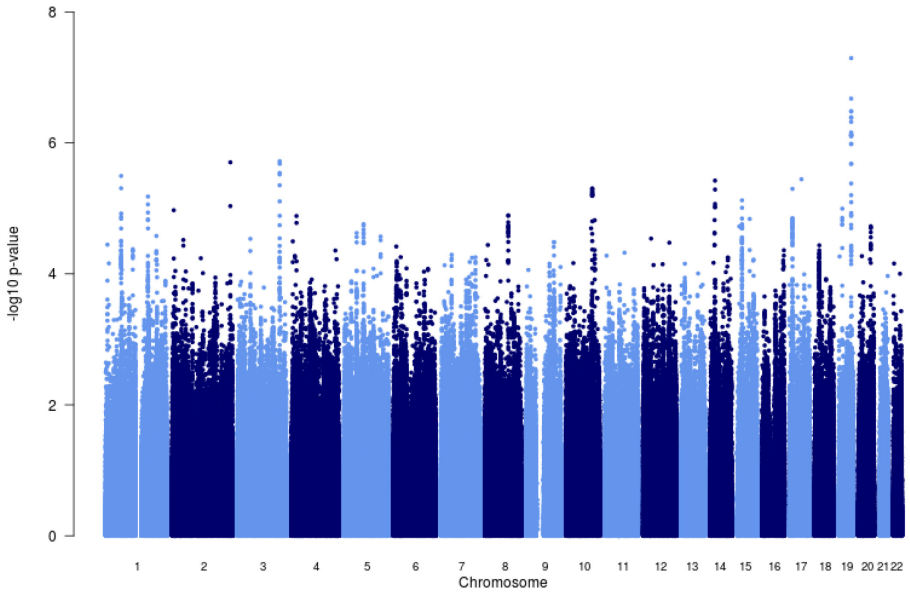


*Fructosamine is natural log transformed

Supplemental Figure 3.12. Regional association plot of rs2438321 and fructosamine among blacks.

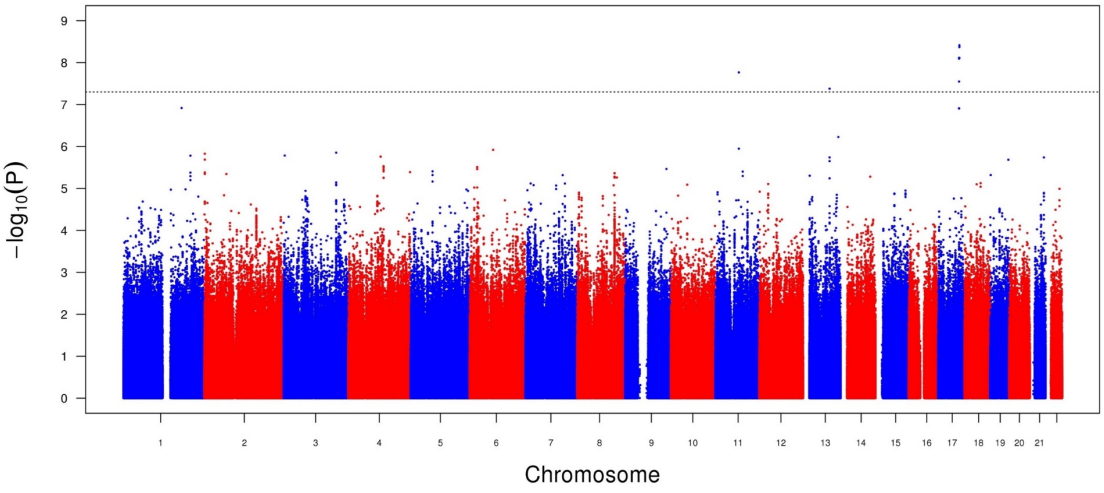


Supplemental Figure 3.13. Manhattan plot for meta-analysis of black and white ARIC participants for fructosamine (N=9,751)



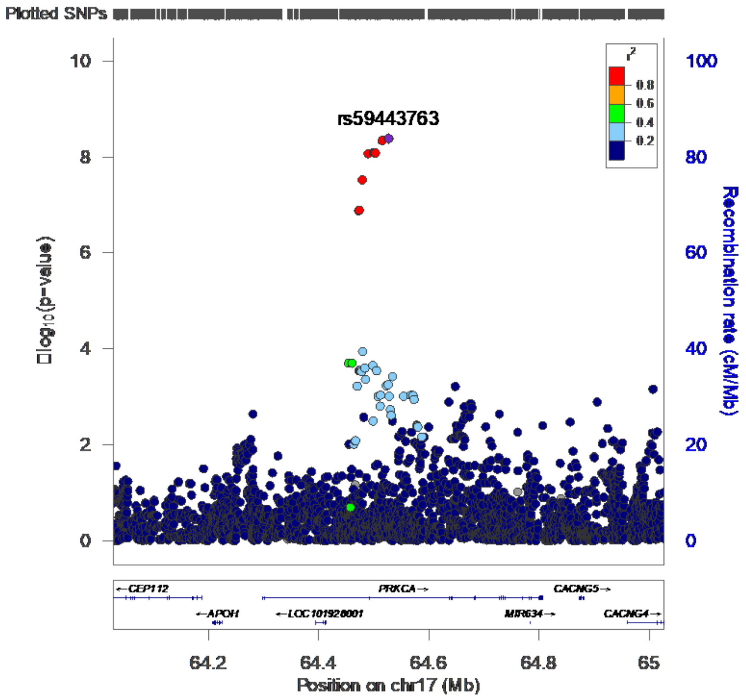
*Fructosamine is natural log transformed

Supplemental Figure 3.14. Manhattan plot for GWAS of percent glycated albumin* in blacks (N=2,104).



*Percent glycated albumin is natural log transformed.

Supplemental Figure 3.15. Regional association plot of rs59443763 and percent glycated albumin in blacks.



Chapter 4: Multivariate phenotype analysis of hyperglycemia biomarkers, fructosamine and glycated albumin

Authors:

Stephanie J. Loomis[1], Elizabeth Selvin[1, 2], Adrienne Tin[1], Josef Coresh[1, 2], Baolin Wu[3], Priya Duggal, [1] Debashree Ray[1]

[1]Department of Epidemiology, The Johns Hopkins University Bloomberg School of Public Health, Baltimore MD

[2]Welch Center for Prevention, Epidemiology, & Clinical Research, The Johns Hopkins University, Baltimore MD

[3]Department of Biostatistics, University of Minnesota, Minneapolis, MN

4.1 Abstract

Fructosamine and glycated albumin are nontraditional type 2 diabetes biomarkers that can be used for diabetes diagnosis and management. These biomarkers are under genetic control, and a previous genome-wide association study in the Atherosclerosis Risk in Communities (ARIC) Study identified several variants associated with fructosamine and glycated albumin. To take advantage of the correlation between these two biomarkers and increase power over the single phenotype analysis to detect additional variants associated with fructosamine and glycated albumin, we implemented a multivariate phenotype analysis using the program Unified Score-based Association Test (USAT). We jointly analyzed fructosamine and glycated albumin among European American (N=7,395) and African American (N=2,309) participants in ARIC using both genotype and exome sequencing data. Among individuals of European ancestry, variants in the *UGT1A* region were associated with a joint fructosamine-glycated albumin phenotype (top SNP rs887829, $p=3.18x10^{-8}$) that was not identified in the single phenotype analyses. No additional loci were identified in African ancestry individuals using this multivariate analysis. The use of a multivariate phenotype identified a new locus, *UGT1A* for fructosamine and glycated albumin, and further research is warranted to understand its role.

4.2 Introduction

Type 2 diabetes is defined by hyperglycemia, or elevated blood glucose concentrations. Hyperglycemia is commonly measured by fasting glucose and HbA1c, but due to limitations of these biomarkers, nontraditional biomarkers such as fructosamine and glycated albumin have been proposed.[1-7] Fructosamine, which is glucose bound to total serum protein, and glycated albumin, which is glucose bound to the most prevalent serum protein, albumin, reflect average blood glucose over the previous 2-3 weeks.[8]

It is well established that genetics play a role in type 2 diabetes and traditional measures of hyperglycemia. Large studies evaluating fasting glucose and HbA1c (sample sizes up to 133,000-160,000) have identified many variants associated with these measures (fasting glucose (n=36) and HbA1c (n=60)).[9,10] Our recent genome-wide association study (GWAS) also identified several genetic variants associated with fructosamine and glycated albumin: a known type 2 diabetes variant in *GCKR* associated with glycated albumin and a variant in *RCN3* associated with fructosamine, possibly in a nonglycemic manner (Chapter 3). However, these variants only explain 0.3 to 0.6% of the variance of glycated albumin (h^2=0.45) and fructosamine (h^2=0.44), respectively, among people of European ancestry (Chapter 2,3). This may be due to the limited sample sizes of this study (N=7,229).

One way to achieve additional power is through multivariate phenotype analysis,[11,12] which aims to jointly analyze multiple phenotypes (e.g., multiple biomarkers of hyperglycemia) by taking into account the correlation structure of the phenotypes. This phenotypic correlation structure is ignored in univariate analyses, which is

equivalent to discarding this information. We used the method Unified Score-based Association Test (USAT)[13] to evaluate the association between genetic variants and biomarkers of hyperglycemia in the Atherosclerosis Risk in Communities (ARIC) Study. USAT is a data-adaptive test of association of multiple continuous phenotypes with a single genetic variant. We performed multivariate phenotype analysis using both GWAS data, which captures common variants, and exome sequencing data, which captures rare coding variants, to identify additional variants associated with fructosamine and glycated albumin.

4.3 Methods

Study Population

The ARIC study is a longitudinal cohort study initiated in 1987 to examine cardiovascular disease risk factors in 15,792 middle-aged adults from four study sites: Jackson, Mississippi; Forsyth, North Carolina; Washington County, Maryland; and suburban Minneapolis, Minnesota.[14] Participants attended a baseline visit in 1987-1989 with six subsequent visits and a seventh currently ongoing. Study participants provided informed consent and protocols were approved by relevant institutional review boards.

This analysis was limited to individuals who had genotyping or exome sequencing data available. Of the 14,348 individuals who attended ARIC visit 2, individuals were excluded if they had diagnosed diabetes (N=1,256), missing diabetes status (N=4), or missing fructosamine or glycated albumin measures (N=929; **Supplemental Figure 4.1**). A total of 9,411 (N=7,395 European ancestry individuals and 2,016 African ancestry individuals) were genotyped and passed quality control, and a

total of 8,899 individuals (N=6,590 European ancestry individuals, 2,309 African ancestry individuals) were exome sequenced and passed quality control.

Genotyping

Samples were genotyped on the Affymetrix 6.0 array and imputed to 1000 Genomes Phase I (March 2012) using IMPUTE2. Poor quality samples were excluded due to sex mismatches, genetic outliers, failed concordance with Taqman genotypes, first-degree relatedness with another study member and missingness>98%. SNPs with missingness>5%, Hardy-Weinberg Equilibrium (HWE)<0.00001, or low minor allele frequency (MAF) <0.005 were excluded. Indels, SNPs with duplicated basepair positions, imputation quality info score<0.8 and minor allele frequency (MAF)<0.05 were also excluded. Imputed dosages were converted to hardcalls (A,T,G,C) for analysis in USAT. The final dataset included 6,792,306 SNPs in the European ancestry sample and 5,313,666 SNPs in the African ancestry sample.

Exome sequencing and Quality Control

DNA was extracted from blood collected at visit 1 from ARIC participants. All sequencing was done as part of the CHARGE consortium at the Baylor College of Medicine Human Genome Sequencing Center (HGSC). Samples were sequenced on the Illumina HiSeq 2000 or 2500 platform (San Diego, CA), and quality control measures implemented. Single nucleotide variants (SNVs) were excluded if they met any of the following criteria: posterior probability<0.95, variant read count <3, variant read ratio <0.25 or >0.75, strand bias >99% in single direction, total convergence <10 fold for SNVs (<30x for indels), outside exon capture regions, monomorphic variant, missing rate >20%, mappability score <0.8, mean depth coverage >500 fold, Hardy Weinberg

Equilibrium $p<5x10^{-6}$ in ancestry-specific groups. Samples were excluded if they had >20% missing data or beyond 6 standard deviations from the mean read depth, singleton count, heterozygote to homozygote ratio, or transition to transversion (Ti/Tv) ratio. The post quality control sample included 7,810 European Ancestry individuals and 3,180 African Ancestry individuals with genetic data on 2,556,859 SNVs and 76,133 indels. In addition, we further excluded variants with less than 10 copies of the minor allele (MAF<0.008 (10/(6,590*2)) for the European ancestry sample and MAF<0.002 (10/(2,309*2) for the African ancestry sample).

Hyperglycemia biomarkers

All biomarkers were measured from serum samples collected at visit 2 (1990-1992). Samples were stored at -70°C and fructosamine (Roche Diagnostics, Indianapolis IN, USA) and glycated albumin (GA-L Asashi Kasei Pharma Corporation, Tokyo, Japan) were measured in 2012-2013 using a Roche Modular P800 system.

Statistical Analysis

Values for all biomarkers were natural log transformed for statistical analyses to account for skewed distributions. We calculated residuals based on models controlling for age, sex, ARIC study site and the top 10 principal components, and used the residuals as phenotypes in all analyses.

Multivariate analysis

We implemented the program USAT (v1.21, https://github.com/RayDebashree/USAT), which has high power to detect association in a variety of variant-phenotype association settings that is lacking in other multivariate methods.[13,15] For a given genetic variant, USAT gives weights to two types of

multivariate phenotype methods: MANOVA (based on a multivariate linear regression model) and a marginal score test (based on univariate regression models). It then adaptively selects the weighted test with minimum p-value and computes an approximate asymptotic p-value of association.[13,15]

We ran USAT using an additive genetic model testing one variant at a time separately by ancestry. We evaluated common variants using genotype data, and rare, coding variants using exome sequencing data combining the phenotypes fructosamine and glycated albumin.

Univariate analysis

To determine the association between the variants and the biomarkers individually using genotype data, we performed linear regression analysis of all variants using an additive genetic model on biomarker residuals using PLINK v1.9 separately by ancestry.[16] We accounted for multiple testing of the two phenotypes by using a significance threshold of $2.5 \times 10^{-8} = 5 \times 10^{-8}/(2 \text{ biomarkers})$.

We also evaluated the association of exome sequencing variants with fructosamine and glycated albumin, separately by ancestry using the R program SeqMeta. We used an additive genetic model and the biomarker residuals. We accounted for multiple comparisons by using a Bonferroni corrected significance threshold of ($2.1 \times 10^{-7} = 0.05/(121{,}052 \text{ variants} * 2 \text{ biomarkers})$ for European ancestry, $1.4 \times 10^{-7} = 0.05/(175{,}583 \text{ variants} * 2 \text{ biomarkers})$ for African ancestry.

4.4 Results

Study characteristics

This analysis included European ancestry individuals: 7,395 with genotype data and 6,590 with exome sequencing data, and African ancestry individuals: 2,016 with genotype data and 2,309 with sequencing data. Over half were female, and the mean age was 56-57 years. Values for all glycemic biomarkers were higher among African ancestry individuals than European ancestry individuals (**Supplemental Table 4.1**). Correlations among biomarkers were moderate (r=0.52 between HbA1c and fructosamine) to strong (r=0.79 for fructosamine and glycated albumin) and were similar in the genotyped and exome sequenced samples (**Supplemental Table 4.2**).

Multivariate phenotype analysis

Among European ancestry individuals, the joint analysis using genotype data of fructosamine and glycated albumin identified variants on chromosome 2 in the *UGT1A* region (top SNP rs887829, $p=3.18x10^{-8}$; **Table 4.1**) that were not significant in the single phenotype analysis for either fructosamine ($p=6.04x10^{-6}$) or glycated albumin (p=0.72). Using the exome sequencing data, no variants were significantly associated with the joint fructosamine-glycated albumin phenotype (**Table 4.2**), however variants in LD ($r^2=0.75$) with the *UGT1A1* variants identified with the genotype data neared significance (rs1105880, $p=1.83x10^{-5}$) in the exome multivariate analysis.

The variant in *GCKR* significantly associated with glycated albumin in our univariate GWAS using imputed data (rs1260326) approached significance in both the genotype ($p=8.60x10^{-8}$) and exome sequencing ($p=1.61x10^{-5}$) multivariate analysis (**Tables 4.1, 4.2**). The variant associated with fructosamine in our previous GWAS (rs34459162) approached significance in the exome sequencing multivariate analysis ($p=2.56x10^{-7}$, **Table 4.2**).

Among African ancestry individuals, genotype multivariate analysis identified two regions on chromosome 11: a variant in the *ARAP1/STARD10* ($p=1.88x10^{-8}$; **Table 4.3**) and an intergenic variant (rs2438321, $p=3.06x10^{-8}$), as well a region on chromosome 17 in *PRKCA* ($p\geq1.88x10^{-8}$) associated with the joint fructosamine-glycated albumin phenotype. The rs2438321 SNP was associated with fructosamine ($p=2.65x10^{-9}$) and the *ARAP1/STARD10* ($p=2.03x10^{-8}$) and *PRKCA* variants ($p\geq3.78x10^{-9}$) were associated with glycated albumin in the univariate GWAS using imputed data. Multivariate analysis using sequencing data did not produce any variants significantly associated with the fructosamine-glycated albumin joint phenotype.

Single phenotype analysis

Among European ancestry individuals, no variants were genome-wide significantly associated with fructosamine or glycated albumin using the genotyped data (**Supplemental Table 4.3**). For fructosamine, the significant SNP in our previous GWAS in *RCN3*, rs34459162 neared significance ($p=1.96x10^{-6}$) and was in linkage disequilibrium (LD) with the most significant SNP in this analysis (rs111981233, $r^2=0.67$, $p=1.75x10^{-7}$). For glycated albumin, the significant SNP in our previous GWAS in *GCKR*, rs1260326, also approached significance ($p=6.22x10^{-8}$). One variant in the European ancestry sample reached exome-wide significance in the univariate analysis (rs34459162, $p=2.19x10^{-7}$; **Supplemental Table 4.4**), which was also identified by our previous GWAS. One variant was associated with glycated albumin among European ancestry individuals (rs184161698, $1.43x10^{-7}$; **Supplemental Table 4.4**).

In African ancestry individuals, variants in an intergenic region on chromosome 11 including the SNP from our original GWAS, rs2438321 ($p=2.65x10^{-9}$) and a variant

on chromosome 2 (rs10490265, $p=1.59x10^{-8}$) were associated with fructosamine using the genotype data, (**Supplemental Table 4.4**). A region on chromosome 17 (top SNP: rs74617220, $p=3.78x10^{-9}$) which included the SNP from our previous GWAS near *PRKCA* and a variant on chromosome 11 ($p=2.03x10^{-8}$) was associated with glycated albumin. Using the exome sequencing data, a variant in *COL19A1* (rs140658141, $p=1.21x10^{-7}$; **Supplemental Table 4.6**) was associated with fructosamine, and two variants were associated with glycated albumin (rs115081997, $p=6.03x10^{-8}$; rs149541658, $p=1.12x10^{-7}$; **Supplemental Table 4.6**).

4.5 Discussion

Among European ancestry individuals, the multivariate phenotype analysis of fructosamine and glycated albumin using common and low-frequency variants identified variants in the *UGT1A* region that was not significant in the univariate analysis of fructosamine or glycated albumin. Exome sequencing analysis also identified variants in this same region approaching significance. The *UGT1A* region is a complex of alternatively spliced genes including *UGT1A11, 1A3, 1A4, 1A5, 1A6, 1A7, 1A8, 1A9* and *1A10* (RefSeq: https://www.ncbi.nlm.nih.gov/nuccore/NM_000463.2). These genes are involved in the glucuronidation of bilirubin (a product of heme catabolism), which creates water-soluble bilirubin. *UGT1A* variants cause the hereditary unconjugated hyperbilirubinemias Crigler-Najjar syndromes and Gilbert syndrome.[17] In addition, variants in *UGT1A* are associated with bilirubin levels among individuals without these syndromes.[18-20] Moderately elevated bilirubin is associated with a decreased risk of diabetes and cardiovascular disease.[21-23] In addition, bilirubin can also bind to albumin.[24]

In the single phenotype analysis, the variants in *UGT1A* show suggestive significance with fructosamine, measured as a concentration that does not account for total serum protein. In contrast, variants in *UGT1A* were not associated with glycated albumin, which is expressed as the percent, accounting for total serum albumin. Because the multivariate phenotype results are likely driven by the fructosamine association, and fructosamine levels are more affected by serum albumin levels, it may be that the multivariate phenotype association is impacted by an albumin-related pathway rather than a diabetes-related pathway. If variants in *UGT1A* affect bilirubin homeostasis, this could alter the amount of albumin bound to bilirubin, which would then impact the amount of albumin available to be glycated, thus impacting fructosamine levels.

The joint analyses using genotype and exome sequencing data support the findings from our previous univariate GWAS using imputed data, identifying all of the same regions as the GWAS: *GCKR, RCN3, PRKCA* and an intergenic region on chromosome 11. The minor differences in the single phenotype analysis presented here from our previous GWAS are likely due to the slightly different sample (this analysis was restricted to individuals who had complete data on all biomarkers), the differences in data (imputed dosages vs imputed data converted to hard calls), as well as the program used for analysis (SNPTEST vs PLINK). The variants associated with fructosamine and glycated albumin in European ancestry individuals in our original GWAS were very close to genome-wide significance in this analysis, and the variants found in African ancestry individuals in the original GWAS were also found to be associated in this analysis.

All of the significant associations from the joint modeling of fructosamine and glycated albumin using the genotyping data in the African ancestry sample were previously identified in the univariate phenotype analysis of fructosamine or glycated albumin. Joint analysis using the exome sequencing data did not result in significant associations in either the European ancestry or African ancestry sample.

Multivariate phenotype analysis successfully demonstrated increased power to detect genetic associations with fructosamine and glycated albumin among European ancestry individuals, by identifying variants in *UGT1A1* not significantly associated in single phenotype analysis. This is an important method, particularly for phenotypes like fructosamine and glycated albumin that are not regularly collected in epidemiologic studies, impeding the ability to increase power by increasing sample size.

4.6 Tables and Figures

Table 4.1. Multivariate phenotype analysis results using genotype data among European ancestry individuals.

SNP	Chr	A1	Gene	USAT P-value[1]	Fructosamine P-value[1]	Glycated albumin P-value[1]
rs887829	2	T	UGT1A1[2]	**3.18E-08**	6.04E-06	0.72
rs4148325	2	T	UGT1A1	**3.52E-08**	9.72E-06	0.83
rs4148324	2	G	UGT1A1	**4.83E-08**	1.18E-05	0.83
rs6742078	2	T	UGT1A1	5.56E-08	1.20E-05	0.82
rs59774409	19	T	FCGRT	5.83E-08	2.75E-07	2.50E-07
rs111981233	19	G	FCGRT	7.87E-08	1.75E-07	5.38E-07
rs1260326	2	T	GCKR	8.60E-08	0.02	6.22E-08
rs111741722	2	G	UGT1A1	9.62E-08	1.48E-05	0.81
rs780093	2	T	GCKR	1.33E-07	0.03	1.14E-07
rs780094	2	T	GCKR	2.56E-07	0.04	2.13E-07

[1]Significant results are in bold. Significance threshold: $p<5x10^{-8}$ for multivariate results, $5x10^{-8}/(2 \text{ biomarkers})=2.5x10^{-8}$ for univariate results.
[2]This region includes multiple alternatively spliced genes including *UGT1A10/UGT1A1/UGT1A3/UGT1A4/UGT1A5.*

Table 4.2. Multivariate phenotype analysis results using exome sequencing data among European ancestry individuals.

SNP	Chr	Gene	A1	RAF[1]	USAT P-value[2]	Fructosamine P-value[2]	Glycated Albumin P-value[2]
rs34459162	19	RCN3	T	0.07	2.56E-07	2.19E-07	**7.56E-07**
rs34654230	19	RCN3	C	0.07	6.96E-07	5.14E-07	2.05E-06
rs3187346	19	RCN3	T	0.07	1.54E-06	1.49E-06	1.70E-06
rs1129459	19	RCN3	C	0.07	1.71E-06	1.18E-06	4.34E-06
rs116604830	12	CCDC63	G	0.02	1.99E-06	9.15E-07	0.0005
rs28929474	14	SERPINA1	C	0.02	5.71E-06	0.92	0.0003
rs8105626	19	RCN3	T	0.07	8.84E-06	5.77E-06	2.19E-05
rs2878342	19	FCGRT	C	0.08	1.28E-05	1.44E-05	5.08E-06
rs1260326	2	GCKR	T	0.59	1.61E-05	0.02	2.94E-06
rs1105880	2	UGT1A6	A	0.35	1.83E-05	9.96E-05	0.56

[1]RAF=risk allele frequency of A1

[2]Significant results are in bold. Significance threshold: 0.05/(121,052 variants * 2 biomarkers)=4.1×10^{-7} for multivariate results, 0.05/(121,052 variants * 2 biomarkers)= 2.1×10^{-7} for univariate results.

Table 4.3. Multivariate phenotype analysis results using genotype data among African ancestry individuals.

SNP	Chr	A1	Gene	USAT P-value[1]	Fructosamine P-value[1]	Glycated albumin P-value[1]
rs116714277	11	T	STARD10	**1.83E-08**	1.19E-07	**2.03E-08**
rs59443763	17	C	PRKCA	**1.88E-08**	1.04E-06	**3.79E-09**
rs59337138	17	T	PRKCA	**2.15E-08**	1.41E-06	**3.78E-09**
rs74617220	17	T	PRKCA	**2.15E-08**	1.41E-06	**3.78E-09**
rs2438321	11	G	intergenic	**3.06E-08**	**2.65E-09**	1.52E-05
rs58597349	17	T	PRKCA	**3.78E-08**	1.88E-06	**7.57E-09**
rs58769090	17	T	PRKCA	**3.78E-08**	1.88E-06	**7.57E-09**
rs73336540	17	C	PRKCA	**3.78E-08**	1.88E-06	**7.57E-09**
rs73336524	17	G	PRKCA	6.11E-08	2.42E-06	**1.35E-08**
rs1348516	11	G	intergenic	6.21E-08	**5.60E-09**	2.37E-05

[1]Significant results are in bold. Significance threshold: $p<5x10^{-8}$ for multivariate results, $5x10^{-8}/(2 \text{ biomarkers})=2.5x10^{-8}$ for univariate results.

Table 4.4. Multivariate phenotype analysis results using exome sequencing data among African ancestry individuals.

SNP	Chr	Nearest Gene	A1	RAF[1]	USAT P-value[2]	Fructosamine P-value[2]	Glycated Albumin P-value[2]
rs115081997	2	DNAH7	A	0.005	7.77E-07	4.86E-05	**6.03E-08**
rs149541658	19	SLC27A5	G	0.009	1.44E-06	8.45E-05	**1.12E-07**
rs9937169	16	PKD1L2	T	0.052	1.09E-05	1.91E-05	0.08
rs237025	6	SUMO4	G	0.717	1.16E-05	5.12E-06	0.001
rs143807085	2	ITGA4	A	0.010	1.49E-05	3.55E-05	0.13
rs73265846	5	PCDHGB3	C	0.016	3.10E-05	1.43E-05	0.004
rs77605839	11	FCHSD2	A	0.028	3.11E-05	1.62E-05	0.0002
rs16925431	9	DMRT1	T	0.076	3.64E-05	1.71E-05	0.002
rs16931976	11	OR51B5	T	0.061	5.69E-05	3.47E-05	6.71E-05
rs2292324	16	NECAB2	C	0.453	6.53E-05	3.11E-05	0.001

[1]RAF=risk allele frequency of A1

[2]Bonferroni corrected significance threshold = 1.4×10^{-7} (0.05/(175,583 variants * 2 biomarkers)), significant results are in bold

[2]Significant results are in bold. Significance threshold: 0.05/175,583 variants $=2.8 \times 10^{-7}$ for multivariate results, 0.05/(175,583 variants * 2 biomarkers)= 1.4×10^{-7} for univariate results.

Supplemental Table 4.1. Demographic and clinical characteristics of study participants[1]

	Genotype sample		Exome sequencing sample	
	European ancestry (N=7,395)	African ancestry (N=2,016)	European ancestry (N=6,590)	African ancestry (N=2,309)
Female	54%	63%	55%	63%
Age	57 (5.7)	56 (5.7)	57 (5.6)	56 (5.7)
Fructosamine (µmol/L)	227 (23)	238 (35)	227 (23)	238 (32)
Glycated albumin (%)	12.6 (1.6)	13.6 (2.6)	12.6 (1.6)	13.6 (2.4)
HbA1c (%)[2]	5.4 (0.5)	5.8 (0.8)	5.4 (0.52)	5.8 (0.86)
Fasting glucose (mg/dL)[2]	103.9 (17.0)	109.2 (24.8)	104 (17)	109 (26)

[1]Continuous variables shown as mean (SD)
[2]HbA1c genotyped sample European ancestry N=7,315, African ancestry N=1,976; exome sequenced sample European ancestry N=6,524, African ancestry N=2,267. Fasting glucose genotyped sample European ancestry N=7,294, African ancestry N=1,941; exome sequenced sample European ancestry N=6,502, African ancestry N=2,223.

Supplemental Table 4.2. Pearson's correlations among biomarkers (genotyped sample N=9,411, exome sequenced sample N=8,621).

	Fructosamine (μmol/L)	Glycated Albumin (%)	HbA1c (%)	Fasting Glucose (mg/dL)
Fructosamine (μmol/L)	1			
Glycated Albumin (%)	0.79, 0.80	1		
HbA1c (%)	0.52, 0.55	0.61, 0.64	1	
Fasting Glucose (mg/dL)	0.58, 0.60	0.61, 0.65	0.72, 0.74	1

Supplemental Table 4.3. Top single variant results for European ancestry individuals, genotyping. [1]

Fructosamine (µmol/L)						
Chr	**SNP**	**BP**	**Nearest Gene**	**A1**	**Beta**	**P-Value**
19	rs111981233	50016479	FCGRT	G	-0.02	1.75E-07
19	rs59774409	50016748	FCGRT	T	-0.02	2.75E-07
19	rs4075250	50057854	intergenic	A	-0.01	4.04E-07
19	rs4594362	50059982	NOSIP	G	-0.01	5.14E-07
1	rs17316247	63258549	ATG4C	C	0.01	1.66E-06
19	rs111508578	50027533	FCGRT	T	-0.02	1.81E-06
19	rs73586516	50046924	RCN3	G	-0.02	1.89E-06
19	rs1316885	50010907	intergenic	C	-0.01	1.93E-06
19	rs113176985	50008951	intergenic	T	-0.01	1.99E-06
19	rs80094695	50047487	intergenic	C	-0.02	2.05E-06
Glycated Albumin (%)						
Chr	**SNP**	**BP**	**Nearest Gene**	**A1**	**Beta**	**P-Value**
2	rs1260326	27730940	GCKR	T	-0.01	6.22E-08
2	rs780093	27742603	GCKR	T	-0.01	1.14E-07
7	rs10255211	86124483	intergenic	C	-0.01	1.53E-07
2	rs780094	27741237	GCKR	T	-0.01	2.13E-07
19	rs59774409	50016748	FCGRT	T	-0.02	2.50E-07
19	rs111981233	50016479	FCGRT	G	-0.02	5.38E-07
6	rs4897175	126657472	intergenic	G	0.01	5.60E-07
2	rs4665972	27598097	SNX17	T	-0.01	9.34E-07
6	rs576049	126730543	CENPW	G	0.01	1.10E-06
1	rs6663454	63295372	ATG4C	A	0.01	1.32E-06

[1]Significant results are in bold: $5x10^{-8}/(2 \text{ biomarkers})=2.5x10^{-8}$

Supplemental Table 4.4. Top single variant results for European ancestry individuals, exome sequencing. [1]

Fructosamine (µmol/L)							
SNP	**Gene**	**Chr**	**A1/A2**	**P-value**[1]	**MAF**	**Beta**	**SE**
rs34459162	RCN3	19	T/C	2.19E-07	0.071	-3.41	0.66
rs34654230	RCN3	19	C/T	5.14E-07	0.073	-3.25	0.65
rs116604830	CCDC63	12	G/T	9.15E-07	0.017	6.52	1.33
rs1129459	RCN3	19	C/T	1.18E-06	0.073	-3.16	0.65
rs3187346	RCN3	19	T/C	1.49E-06	0.072	-3.14	0.65
rs8105626	RCN3	19	T/C	5.77E-06	0.068	-3.04	0.67
rs137994820	ARAP3	5	T/C	7.41E-06	0.001	26.13	5.83
rs76384862	COMT	22	C/T	8.19E-06	0.001	24.90	5.58
rs184161698	ID1	20	A/T	1.17E-05	0.001	18.95	4.32
rs139240119	SLC25A25	9	G/T	1.20E-05	0.010	7.30	1.67
Glycated Albumin (%)							
SNP	**Gene**	**Chr**	**A1/A2**	**P-value**[1]	**MAF**	**Beta**	**SE**
rs184161698	ID1	20	A/T	**1.43E-07**	0.001	1.54	0.29
rs34459162	RCN3	19	T/C	7.56E-07	0.071	-0.22	0.04
rs202140443	PSMB10	16	T/G	1.30E-06	0.001	1.49	0.31
rs3187346	RCN3	19	T/C	1.70E-06	0.072	-0.21	0.04
rs142744907	COG7	16	C/T	1.97E-06	0.001	1.61	0.34
rs34654230	RCN3	19	C/T	2.05E-06	0.073	-0.21	0.04
rs1260326	GCKR	2	T/C	2.94E-06	0.407	0.11	0.02
rs1129459	RCN3	19	C/T	4.34E-06	0.073	-0.20	0.04
rs2878342	FCGRT	19	C/T	5.08E-06	0.076	-0.20	0.04

[1]Significant results are in bold: 0.05/(121,052 variants * 2 biomarkers)= 2.1×10^{-7}

Supplemental Table 4.5. Top single variant results for African ancestry individuals, genotyping.

Fructosamine (µmol/L)						
Chr	**SNP**	**BP**	**Nearest Gene**	**A1**	**Beta**	**P-Value**[1]
11	rs2438321	98500410	intergenic	G	0.04	**2.65E-09**
11	rs1348516	98495807	intergenic	G	0.04	**5.60E-09**
11	rs2438322	98500409	intergenic	T	0.04	**9.10E-09**
11	rs1348518	98495639	intergenic	G	0.04	**1.07E-08**
2	rs10490265	40979468	LOC101929700	T	0.03	**1.59E-08**
11	rs2512816	98490543	intergenic	T	0.03	2.73E-08
11	rs3018601	98508879	intergenic	A	0.03	4.14E-08
11	rs2512863	98507758	intergenic	T	0.03	5.12E-08
11	rs2512825	98513495	intergenic	A	0.03	7.10E-08
2	rs55814013	40982682	LOC101929700	A	0.03	7.84E-08
Glycated Albumin (%)						
Chr	**SNP**	**BP**	**Nearest Gene**	**A1**	**Beta**	**P-Value**[1]
17	rs74617220	64515083	PRKCA	T	0.05	**3.78E-09**
17	rs59337138	64515907	PRKCA	T	0.05	**3.78E-09**
17	rs59443763	64526988	PRKCA	C	0.05	**3.79E-09**
17	rs73336540	64499524	PRKCA	C	0.05	**7.57E-09**
17	rs58769090	64501060	PRKCA	T	0.05	**7.57E-09**
17	rs58597349	64503204	PRKCA	T	0.05	**7.57E-09**
17	rs73336524	64489878	PRKCA	G	0.05	**1.35E-08**
11	rs116714277	72473447	STARD10	T	0.06	**2.03E-08**
11	rs114715767	72449727	ARAP1	G	0.05	1.09E-07
21	rs67610389	43294820	PRDM15	A	0.03	1.20E-07

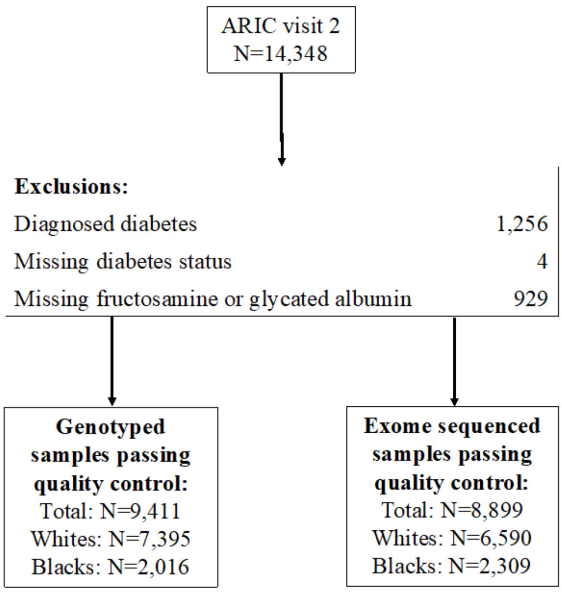
[1]Significant results are in bold: $5x10^{-8}/(2 \text{ biomarkers})=2.5x10^{-8}$

Supplemental Table 4.6. Top single variant results for African ancestry individuals, exome sequencing.

Fructosamine (µmol/L)							
SNP	**Gene**	**Chr**	**A1/A2**	**P-value**[1]	**MAF**	**Beta**	**SE**
rs140658141	COL19A1	6	T/C	**1.21E-07**	0.003	31.39	5.93
rs139579535	COL19A1	6	C/T	4.27E-07	0.005	25.62	5.07
rs142200680	TTC16	9	T/C	1.36E-06	0.002	34.51	7.14
rs139597330	FGL1	8	T/C	2.91E-06	0.005	24.22	5.18
rs1480361	DOCK3	3	T/C	3.02E-06	0.003	-29.66	6.35
rs149282922	SLC13A3	20	C/T	4.04E-06	0.003	28.22	6.12
rs237025	SUMO4	6	G/A	5.12E-06	0.283	3.53	0.77
rs149804703	TDRD7	9	G/A	8.94E-06	0.003	28.15	6.34
rs73265846	PCDHGB3	5	C/T	1.43E-05	0.016	12.12	2.79
rs77605839	FCHSD2	11	A/G	1.62E-05	0.028	9.50	2.20
Glycated Albumin (%)							
SNP	**Gene**	**Chr**	**A1/A2**	**P-value**[1]	**MAF**	**Beta**	**SE**
rs115081997	DNAH7	2	A/G	**6.03E-08**	0.005	1.72	0.32
rs149541658	SLC27A5	19	G/A	**1.12E-07**	0.009	1.35	0.26
rs150448873	RNASE11	14	G/A	9.39E-07	0.002	2.34	0.48
rs139597330	FGL1	8	T/C	5.04E-06	0.005	1.58	0.35
rs184647368	UGT2B4	4	G/A	5.78E-06	0.003	1.99	0.44
rs149804703	TDRD7	9	G/A	6.25E-06	0.003	1.91	0.42
rs375699173	PLEC	8	C/T	7.41E-06	0.002	2.24	0.50
rs201505339	FOSL2	2	A/G	9.96E-06	0.006	1.31	0.30
rs11860488	CES5A	16	G/C	1.40E-05	0.422	0.21	0.05

[1]Significant results are in bold: 0.05/(175,583 variants * 2 biomarkers)= $1.4x10^{-7}$

Supplemental Figure 4.1. Sample exclusions



Chapter 5: Rare variants in *SLC5A10* are associated with serum 1,5-anhydroglucitol (1,5-AG) in the Atherosclerosis Risk in Communities (ARIC) Study

Authors:

Stephanie Loomis[1], Anna Köttgen[1,2], Man Li[1,3], Adrienne Tin[1], Joesf Coresh[1,4], Eric Boerwinkle[5], James Pankow[6], Elizabeth Selvin[1, 4]*, Priya Duggal[1]*

[1]Department of Epidemiology, The Johns Hopkins University Bloomberg School of Public Health, Baltimore MD

[2]Institute of Genetic Epidemiology, Medical Center and Faculty of Medicine - University of Freiburg, Freiburg, Germany

[3]Division of Nephrology and Department of Human Genetics, University of Utah, Salt Lake City, Utah

[4]Welch Center for Prevention, Epidemiology, & Clinical Research, The Johns Hopkins University, Baltimore MD

[5]Department of Epidemiology, The University of Texas Health Science Center at Houston School of Public Health at Houston, Houston, TX

[6]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN

*contributed equally to this work

5.1 Abstract

Introduction: Serum 1,5-anhydroglucitol (1,5-AG) is a measure of hyperglycemic excursions that is available as a clinical test to aid monitoring of glycemic control in persons with diabetes. Recent genome-wide association studies (GWAS) have identified several 1,5-AG associated common variants. Rare genetic variants may have large effects on the concentrations of biomarkers that are in clinical use. The association between rare coding variants and 1,5-AG concentrations measured with a targeted assay of 1,5-AG was investigated in individuals of European and African ancestries.

Methods: Whole exome sequencing association analysis was performed on 1,5-AG among European ancestry (N=6,589) and African ancestry (N=2,309) participants without diagnosed diabetes in the Atherosclerosis Risk in Communities (ARIC) Study. Both single variant and gene-based tests were conducted.

Results: Five variants representing 3 independent signals on chromosome 17 in *SLC5A10* were associated with lower 1,5-AG levels of up to 10.38 µg/mL per allele (1,5-AG range 3.4-32.8 µg/mL). Two of these variants (rs61741107, p=8.85E-56, amino acid change: G>E; rs148178887, p=1.13E-36, amino acid change: N>I) with similar allele frequencies in the European (MAF=0.002 to 0.007) and African (MAF=0.0005 to 0.004) ancestry samples were rare, nonsynonymous, and predicted to be damaging or deleterious by multiple algorithms that predict deleteriousness. Taken together, these three signals explained 6% of the variance in 1,5-AG. Gene-based SKAT-O tests were significant for *SLC5A10* in the European ancestry ($p=5.13x10^{-64}$) and validated in the African ancestry (p=0.006) samples. Significant common and low frequency variants were identified at

five other loci in or near *MUC1, LCT, SI, MGAM*, and *SLC5A1* that also contain common 1,5-AG associated variants.

Conclusions: Several rare, protein-altering variants in individuals of European ancestry were associated with 1,5-AG concentrations and were validated in an African ancestry sample. The large effect sizes for *SLC5A10* variants and similar allele frequencies across populations without diabetes, along with the multiple independent signals are evidence of the important impact of *SLC5A10* on serum 1,5-AG concentrations, and suggest *SLC5A10* may code for an important transporter of 1,5-AG in the kidney. This study helps to characterize the genetic architecture of 1,5-AG, an emerging diabetes biomarker, and may have implications for its clinical interpretation.

5.2 Introduction

1,5-andhydroglucitol (1,5-AG) is an emerging biomarker of glycemic control in type 2 diabetes. 1,5-AG is a monosaccharide consumed in food and maintained at high, constant levels in the blood under normoglycemic conditions through filtration by the kidney and reabsorption into the blood. 1,5-AG is the 1-deoxy form of glucose, and during hyperglycemic conditions (i.e., when glucose exceeds the renal threshold), glucose outcompetes 1,5-AG for reabsorption. This causes 1,5-AG excretion in urine and hence lower levels in blood concentrations.[1] In adults with diabetes, low 1,5-AG concentrations reflect glucose excursions over the previous 2-14 days,[1,2] and are associated with microvascular and macrovascular disease.[3,4,5]

In a recent genome-wide association study (GWAS), we identified seven variants at six loci associated with 1,5-AG among persons of European ancestry without diagnosed diabetes.[6] Two of these variants were also found in a genetic screen of 1,5-AG measured as part of a large non-targeted metabolome panel among Europeans.[7] These variants map in or near genes which are involved in carbohydrate metabolism (*LCT, SI, MGAM, MGAM2*) and glucose transport in the gut and kidney (*SLC5A10, SLC50A1, SLC5A1*).[6] Interestingly, the majority of the variants are not associated with traditional measures of hyperglycemia such as fasting glucose and HbA1c.[8,9]

While array-based analyses such as GWAS often capture common variants in linkage disequilibrium with the putative causal variants, they are not able to assess the impact of rare variants, nor are they designed to identify causal variants. To investigate the association of rare, putatively damaging variants with 1,5-AG, and to further understand the genetic architecture of this biomarker, a whole-exome sequencing

association study of 1,5-AG concentrations was performed in the Atherosclerosis Risk In Communities (ARIC) Study.

5.3 Methods

Study population

The ARIC Study is an ongoing, longitudinal cohort study initiated in 1987, when middle-aged adults were recruited from four communities in the U.S.: Forsyth, North Carolina; suburban Minneapolis, Minnesota; Washington County, Maryland and Jackson, Mississippi. A total of 15,792 individuals attended the initial study visit (1987-1989), and subsequent visits occurred in 1990-1992 (visit 2), 1993-1995 (visit 3), 1996-1998 (visit 4), and 2011-2013 (visit 5), and 2016-2017 (visit 6) with a seventh visit ongoing. The study protocol was approved by the relevant Institutional Review Boards, and all study participants provided written informed consent.[10] For this study, data from individuals who attended visit 2 (N=14,348) was used.

1,5-AG measurement

1,5-AG concentrations were measured using a colometric assay where 1,5-AG is oxidized to hydrogen peroxide (GlycoMark, Winston-Salem, NC) by the Roche Modular P800 system. Serum was collected at visit 2 (1990-1992) and analyzed in 2012-2013. The interassay coefficient of variation was 4.8%.[11]

Exome sequencing

DNA was extracted from blood primarily collected at visit 1. All sequencing was performed as part of the CHARGE Consortium exome sequencing project at the Baylor College of Medicine Human Genome Sequencing Center (HGSC). Samples were bar-

coded, pooled and sequenced using paired-end sequencing, run on the Illumina HiSeq 2000 or 2500 platform (San Diego, CA), and exome capture performed with VCRome 2.1 (NimbleGen, Inc., Madison, WI). Sequence alignment was done using the Burrows-Wheeler alignment[12] tool with the Genome Reference Consortium Human Build 37 reference sequence. Aligned reads were then recalibrated using the Genome ANalysis ToolKit (GATK). Variant calling was done with the Mercury pipeline (https://www.hgsc.bcm.edu/content/mercury) in DNAnexus. VCF files were generated using the Atlas2 suite (Atlas-SNP and Atlas-Indel).

Quality Control

Standard quality control exclusion measures were implemented to ensure accurate, reliable results. Single nucleotide variants (SNVs) were excluded if they met any of the following criteria: posterior probability<0.95, variant read count <3, variant read ratio <0.25 or >0.75, strand bias >99% in single direction, total coverage<10 fold for SNVs (<30x for indels), outside exon capture regions, monomorphic variant, missing rate >20%, mappability score <0.8, mean depth coverage >500 fold, Hardy Weinberg Equilibrium $p<5x10^{-6}$ in ancestry-specific groups. Samples were excluded if they had >20% missing data or fell less than 6 standard deviations (SD) from mean read depth, more than 6 SD for singleton count, outside of 6 SD for heterozygote to homozygote ratio or transition to transversion (Ti/Tv) ratio. After quality control, 2,556,859 SNVs and 76,133 indels remained, and 7,810 European Ancestry individuals and 3,180 African Ancestry individuals remained. Individuals who did not attend visit 2 (N=594), were missing diabetes status (N=2), had diagnosed diabetes (self-reported physician diagnosis or use of diabetes medications; N=875), or missing 1,5-AG data at visit 2 (N=621;

Supplemental Figure 5.1) were also excluded. In total, 6,589 European ancestry samples and 2,309 African ancestry samples were analyzed.

Variant annotation and functional prediction

ANNOVAR8 and dbNSFP v2.0 (https://sites.google.com/site/jpopgen/dbNSFP) were used to annotate variants to genes and functional predictions using the GRCh37 reference sequence and National Center for Biotechnology Information RefSeq. Functional annotation by several metrics predicted if a variant was expected to be damaging (an amino acid change which negatively impacts protein function) or deleterious (a variant which reduces fitness and is subject to purifying selection). SIFT score predicts if an amino acid change is likely to be damaging to protein function based on conservation (i.e., well conserved regions are assumed to be biologically important and thus variants in these regions are more likely to be damaging).[13] A SIFT score <0.05 was considered damaging. Polyphen-2 flags amino acid changes that are predicted to be damaging based on the structure and function of a protein. Polyphen-2 score >0.957 was considered damaging, and a score between 0.453 and 0.965 was considered possibly damaging.[14] GERP predicts substitutions that would have occurred if the region was not under selection and quantifies rate of substitutions that did not occur.[15] A GERP score > 2 was considered deleterious. Finally, CADD aggregates annotations of allelic diversity, functionality, pathogenicity, disease severity, regulatory effects, complex trait associations, and known pathogenic variants into a score.[16] A CADD score >15 was considered deleterious. In addition, the Bravo portal (https://bravo.sph.umich.edu/freeze5/hg38/) was used to obtain TOPMed and 1000 Genomes allele frequencies (Freeze 5, including 463 million variants on 62784

individuals). For each significant variant, the Genotype-Tissue Expression (GTEx) Project was searched for expression quantitative trait loci (eQTLs; https://www.gtexportal.org/home/).

Single-variant tests

Genetic associations with 1,5-AG were analyzed using both single-variant and gene-based tests using the R package SeqMeta. All analyses were run separately by ancestry. 1,5-AG values were winsorized at 1% and 99% to account for long tailed distributions. Single variant analyses were run as linear regressions controlling for age, sex, ARIC study center and significantly associated principal components ($p<0.05$; N=2 for European ancestry, 1 for African ancestry). To ensure our results were not driven by a very small number of individuals, variants with less than 10 copies of the minor allele (minimum MAF=10/(N*2)) were excluded. For the European ancestry sample MAF<0.008 (10/(6589*2)) was used and for the African ancestry sample MAF<0.002 (10/(2309*2)) was used. A Bonferroni correction to calculate a statistical significance threshold as $1.4x10^{-7}$ (0.05/121,052 variants) was used for European ancestry and $2.8x10^{-7}$ (0.05/175,583 variants) was used for African ancestry. Variance explained by individual variants was calculated as the difference between the coefficient of determination from the null model (the association between 1,5-AG and covariates) and a model adjusting for effect of the variant controlling for the same covariates.

Gene-based tests

To augment power for situations where multiple rare variants affect association with a phenotype, the SKAT-O test was run, which aggregates variants into genes and tests for association between genes and phenotypes. SKAT-O combines a burden test, which has

greater power when variants are associated with the phenotype and in the same direction, with SKAT, a kernel based, variance components test, which has greater power when fewer variants are causal or affect risk in both directions. Genes with ≤1 variant per gene were excluded. Variants were not filtered by MAF in the main analysis. Variants other than nonsynonymous, splicing, stop-gain, stop-loss, or frameshift were excluded from the association analyses. Additionally, genes in which all variants used for the burden test together had a cumulative MAF<0.005 were excluded. A Bonferroni correction was used to calculate a significance threshold: 4.0×10^{-6} (0.05/12,504 genes) for European ancestry and 3.3×10^{-5} (0.05/14,499 genes) for African ancestry. Secondary analyses of SKAT and a T1 burden test (where all variants with MAF<0.01 were collapsed into a score for each gene) were also done.

Conditional analyses

To determine if the single-variant results represent independent signals, the top variant (defined as most significant and most deleterious or damaging by GERP, SIFT, Polyphen2 and CADD) was conditioned on for each locus with multiple significant variants. Secondary conditioning analyses were also performed on the most significant variant and the previous GWAS-identified variant. Regional association plots using LocusZoom (http://locuszoom.org/).[17] were created to visualize the region prior to and after conditioning on the top variants.

Variant association with diabetes

To determine if variants significantly associated with 1,5-AG also impact diabetes, the association between these variants and prevalent diabetes status was evaluated. This analysis was performed for both diagnosed diabetes (self-reported physician diagnosis or

use of diabetes medications) and the combination of diagnosed diabetes and undiagnosed diabetes (defined as fasting glucose ≥126 mg/dL if fasting for ≥8 hours or non-fasting glucose ≥200 mg/dL).

5.4 Results

Study population

There were 6,589 individuals in the European ancestry sample and 2,309 individuals in the African ancestry sample. In both groups, over half were female, and mean age was 56 to 57 years old. 1,5-AG levels were lower, and fructosamine, glycated albumin, fasting glucose and HbA1c were higher in the African ancestry sample as compared to the European ancestry sample. Study population characteristics are detailed in **Supplemental Table 5.1**.

Single variant and gene-based analyses

In the European ancestry sample, 15 variants reached exome-wide significance for association with 1,5-AG in single variant testing (**Table 5.1**). These variants are located in 6 loci on chromosomes 1, 2, 3, 7, 17, and 22, all of which were also identified in our previous GWAS of 1,5-AG concentrations.[6] None of the African ancestry single variant or gene-based results were statistically significant.

Rare, deleterious variants on chromosome 17

Four rare (MAF<0.007) and one low frequency (MAF=0.04) variants in the region of two overlapping genes on chromosome 17, *SLC5A10* and *FAM83G* (**Table 5.1**) were associated with 1,5-AG. Of the four rare variants, two were nonsynonymous to *SLC5A10* and were highly significant (rs148178887, $p=1.13x10^{-36}$ and rs61741107, $p=8.85x10^{-56}$).

In addition, the effect sizes of these variants were large, approximately 10 µg/mL per risk allele, and explained 1.71% and 2.95% of the variance in 1,5-AG concentrations, respectively. The other two rare variants were intronic to *SLC5A10* but nonsynonymous to *FAM83G* (rs200038747, $p=1.69x10^{-13}$ and rs201046878, $p=1.96x10^{-29}$). The low frequency variant (rs117355297, $p=3.85x10^{-26}$) was also found in the GWAS[6] and was synonymous to *SLC5A10*. All four nonsynonymous variants were predicted to be damaging or deleterious by the prediction programs GERP, Polyphen-2, SIFT and CADD. In addition, the nonsynonymous variants resulted in amino acid changes which altered polarity and acidity (for example, rs61741107 resulted in a change from nonpolar glycine to acidic glutamic acid, and rs148178887 resulted in a change from polar asparagine to nonpoplar isoleucine). Four variants were also nominally ($p<0.05$) associated with 1,5-AG in African ancestry individuals. The gene-based SKAT-O test showed significance for *SLC5A10* and *FAM83G* (**Table 5.2**). Secondary analyses of separate SKAT ($p=2.8x10^{-55}$), T1 burden ($p=2.5x10^{-114}$) and SKAT-O restricting variants to MAF<0.05 ($p=5.1x10^{-64}$) tests also showed strong significance for *SLC5A10* **(Supplemental tables 5.2-4).**

To determine if the variants in this region were representing one signal in linkage disequilibrium or several independent signals, the nonsynonymous variants were conditioned on (**Figure 5.1**). After conditioning on rs61741107, the variants rs148178887, rs201046878, rs200038747 and rs117355297 remained significant. After additionally conditioning on rs148178887, only rs117355297 remained significant. Further conditioning on the synonymous variant, rs117355297, produced no significant variants in this region. Secondary conditioning on the previously identified GWAS

variant (rs117355297) showed the rare variants remained significant ($p<2.7x10^{-14}$). This suggests that these variants represent three independent, significant loci, which together explain 6% of the variance in 1,5-AG (**Table 5.1**). GTEx did not show eQTLs for any of the chromosome 17 variants in diabetes-relevant tissue such as the kidney, liver or pancreas.

To further explore the rare variants in chromosome 17, they were evaluated for an association with diabetes. Of the variants representing the three significant signals on chromosome 17, none were significantly associated with diagnosed or diagnosed+undiagnosed diabetes status in either European or African ancestry samples (**Supplemental Table 5.5**). In addition, mean 1,5-AG levels differed substantially between individuals with and without the chromosome 17 variants, while the mean values of other glycemic biomarkers did not. (**Figure 5.2, Supplemental Figure 5.2**). No individuals were homozygous for rs61741107 or rs148178887, but eight people were homozygous for rs117355297 (**Supplemental Figure 5.3**). Four individuals had both rs61741107 and rs117355297 (mean 1,5-AG=2.8 μg/mL, SD=1.3 μg/mL), 23 had both rs148178887 and rs117355297 (mean 1,5-AG=9.5 μg/mL, SD=3.8 μg/mL), and two people were heterozygous for rs148178887 and homozygous for rs117355297 (mean 1,5-AG=2.8 μg/mL, SD=0 μg/mL; **Figure 5.2**).

Additional regions of interest

One common variant in *MUC1* on chromosome 1 in linkage disequilibrium (LD) with a variant identified in our GWAS (rs9330264, r^2=0.5) was associated with 1,5-AG, but the gene-based test was not significant. Neither the single variant or gene-based test was validated in the African ancestry sample. GTEx indicated possible eQTLs in

diabetes-related tissues (liver: *GBAP1* $p=1.6x10^{-11}$, *THBS3* $p=2.5x10^{-6}$; pancreas: *GBAP1* $p=3.3x10^{-25}$, *THBS3* $p=5.2x10^{-11}$, *GBA* $p=6.3x10^{-8}$).

Five common variants in genes *LCT, RAB3GAP1, R3HDM1,* and *UBXN4* were associated on chromosome 2 across a large region spanning 0.7 Mb. Three of these variants (rs961360, rs1050115, rs2304371) were in LD with the GWAS index variant, rs182549 (r^2=0.27 to 0.35 in 1000 genomes phase3v5 European population). Two of the five variants were nonsynonymous, one of which (rs961360) was predicted to be possibly damaging by Polyphen-2. The remaining three variants were synonymous and one was also associated in African ancestry individuals (rs1050115, p=0.01). Conditional analysis on the top nonsynonymous variants revealed two distinct signals in this region (**Supplemental Table 5.6**). The GWAS index variant was not present in this dataset and hence could not be conditioned on. None of the genes in this region were associated with 1,5-AG in the gene-based test.

One common variant on chromosome 3 in *SI* was associated with 1,5-AG in European ancestry individuals. This variant is in near perfect LD with the GWAS index variant, rs9825346 (r^2=0.98). It is a nonsynonymous variant, but was not predicted to be damaging or deleterious by any of the prediction programs and was not significant in African ancestry individuals.

One low-frequency (MAF=0.01) variant on chromosome 7 in *MGAM* was associated with 1,5-AG. This variant is not in LD with the GWAS index variant. It is nonsynonymous and predicted to be damaging or deleterious by GERP, Polyphen-2, SIFT and CADD, but was not significant in African ancestry individuals. *MGAM* was associated with 1,5-AG in the gene-based test.

Finally, two common variants in *SLC5A1* were associated in this region. Both variants were in near perfect LD with each other and the GWAS index variant, rs117086479 (r^2=0.98 to 1). One variant was nonsynonymous (rs17683011) and the other was synonymous (rs17683448). Neither had evidence for deleteriousness by any measure, and neither variant was associated with 1,5-AG in African ancestry individuals. *SLC5A1* was significant in the gene-based test.

5.5 Discussion

In this exome sequencing analysis, 15 variants were significantly associated with 1,5-AG among people of European ancestry without diabetes, and four of these variants in two loci were validated in a sample of African ancestry individuals. In addition, 4 genes were associated with 1,5-AG among individuals of European ancestry, of which one (*SLC5A10*) validated in the African ancestry sample.

Both single variant and gene-based tests identified a region on chromosome 17 in or near *SLC5A10* and the overlapping gene, *FAM83G*. *SLC5A10* is a glucose transporter exclusively expressed in the kidney,[18] and is not known to also transport 1,5-AG. Our results, however, suggest *SLC5A10* may be an important transporter of 1,5-AG. Conditional analysis identified multiple distinct signals in this locus. Two of the variants identified (rs61741107 and rs148178887) were also found in a whole genome sequencing analysis of a metabolome panel,[19] adding further evidence to the importance of this region in influencing 1,5-AG levels. The effect sizes of most of the *SLC5A10* variants are large. Given the distribution of 1,5-AG in this sample (in European ancestry: 3.4 to 38.2 winsorized) having just one copy of the rs61741107 or rs148178887 allele would

result in a lowering of 1,5-AG by 10 µg/mL on average. Although no individuals in this dataset are homozygous for rs61741107 (1000 genomes European ancestry MAF = 0.002) or rs148178887 (1000 genomes European ancestry MAF=0.005), the allele frequencies indicate that such individuals do exist in the population, and would have lowering of 1,5-AG levels of over 20 µg/mL on average. In addition, these effect sizes and allele frequencies were similar across ancestries. The smaller p value for the T1 gene-based test as compared to the SKAT test indicates that these variants impact 1,5-AG levels in the same direction. The similar p-value for SKAT-O when restricting variants to MAF<0.05 indicates that the relevant variants are low-frequency and rare.

Many of the variants in this region are predicted to be damaging or deleterious by multiple programs. *SLC5A10* partially overlaps with *FAM83G*, which is expressed in the skin and esophagus (eQTLs; https://www.gtexportal.org/home/)[18] . It is not likely that variants in this region represent diabetes-related factors; neither gene is known to impact diabetes risk, fasting glucose or HbA1c. In addition, variants near *SLC5A10* were not associated with diabetes status, and comparing individuals with and without variants in *SLC5A10*, multiple other measures of hyperglycemia were similar, while mean 1,5-AG differed substantially (**Supplemental Figure 5.2**). Given this evidence, it is likely that rs61741107 and rs148178887 represent putative causal variants for 1,5-AG in this region.

In addition, significant, common variants on chromosomes 1, 2, 3 and 22, and a low frequency variant on chromosome 7 were associated with 1,5-AG concentrations. These loci were all identified by our previous GWAS in the region near *LCT/ UBXN4/ R3HDM1/ RAB3GAP1*. While several of the variants in this region are nonsynonymous, they were mainly not predicted to be damaging or deleterious, and the effect sizes are

relatively small, indicating a potentially more modest impact on 1,5-AG levels. Other regions which were significantly associated among Europeans but not Africans including *MGAM*, *SLC5A1* and *SI*. Further studies are needed to confirm the role of rare variants in these regions in 1,5-AG.

There is currently debate about the utility of 1,5-AG as a useful biomarker of hyperglycemia in adults with diabetes. Prior to widespread use of any clinical test, it is important to identify limitations overall or for specific subpopulations. Warren et al have shown that there is a proportion of individuals for whom 1,5-AG produces "false positive results", i.e., where 1,5-AG concentrations are low while fasting glucose and 2-hour glucose levels are not elevated (manuscript in preparation). [20] Our work shows evidence of a strong genetic impact on 1,5-AG unrelated to diabetes, which may explain some of these findings. The likely nonglycemic genetic impact on 1,5-AG identified in this work is similar to previous findings in HbA1c, for which variants have been identified which impact HbA1c levels, but are not important mechanisms of glucose control.[9,21,22] Our results may have implications for the overall utility of the biomarker independent of genetic characterization.

Exome sequencing has highlighted the role of *SLC5A10* influencing 1,5-AG levels. This study has provided insight into the biology of this biomarker. Although these rare variants impact a smaller number of individuals than common variants, the large effect sizes would likely alter 1,5-AG levels in a sufficient manner to substantially impact its usefulness as a biomarker of hyperglycemia for carriers.

5.6 Tables and Figures

Table 5.1. Significant[1] 1,5-AG (μg/mL) single SNP results in European ancestry sample, with validation in the African ancestry sample.

										European ancestry (N=6,589)				**African ancestry (N=2,309)**			
SNP	Gene	Chr	A1/ A2[2]	Function[3]	Amino acid change	GERP, SIFT, Poly-phen2, CADD Prediction[4]	TOP Med AF	TGP EA Effect AF[5]	TGP AA Effect AF[5]	Effect AF	Beta (SE)	P-value	% Var. Explained	Effect AF	Beta (SE)	P-value	% Var. Explained
rs61741107	*SLC5A10*	17	G/A	**NS**	G>E	**D,D,D,D**	0.004	0.002	0	0.007	-9.31 (0.59)	8.85E-56	2.95	0.0005[7]	-9.17 (3.95)	0.02	0.26
rs148178887	*SLC5A10*	17	A/T	**NS**	N>I	**D,D,D,D**	0.002	0.005	0	0.004	-10.38 (0.82)	1.13E-36	1.71	0.002[7]	-9.93 (2.80)	3.83E-04	0.36
rs201046878	*SLC5A10/ FAM83G*[6]	17	G/A	**NS**/I[5]	R>W	**D,D,D,D**	0.002	0.005	0.002	0.004	-8.33 (0.74)	1.96E-29	1.25	0.002[7]	-9.93 (2.80)	3.83E-04	0.33
rs200038747	*SLC5A10/ FAM83G*[6]	17	C/T	**NS**/I[5]	R>Q	**D,D,D,D**	0.002	0.001	0.005	0.002	-9.04 (1.23)	1.69E-13	0.61	0.004	0.25 (1.25)	0.84	0.09
rs117355297	*SLC5A10*	17	C/T	S	--	T,NA,NA,**D**	0.022	0.05	0.001	0.04	-2.73 (0.26)	3.85E-26	1.37	0.005	-3.34 (1.12)	2.91E-03	0.2
rs4072037	*MUC1*	1	C/T	**NS**	--	T,NA,NA,T	0.60	0.55	0.60	0.54	-0.49 (0.10)	3.74E-07	0.26	0.67	-0.22 (0.18)	0.21	0.1
rs961360	*R3HDM1*	2	A/G	**NS**	M>V	**D**,T,B/**P,**T	0.22	0.23	0.25	0.15	-0.80 (0.14)	7.82E-09	0.32	0.20	-0.41 (0.21)	0.05	0.3
rs10445686	*RAB3GAP1*	2	A/G	**NS**	N>S	**D**,T,B,T	0.14	0.19	0.02	0.13	-0.79 (0.14)	3.59E-08	0.35	0.03	0.21 (0.49)	0.66	0.04
rs2304371	*LCT*	2	G/A	S	--	**D**,NA,NA,T	0.70	0.75	0.41	0.83	0.89 (0.13)	6.74E-12	0.49	0.45	0.39 (0.16)	0.02	0.13
rs3739022	*LCT*	2	G/A	S	--	T,NA, NAT	0.16	0.15	0.22	0.10	-1.07 (0.17)	1.23E-10	0.51	0.21	-0.48 (0.20)	0.02	0.09
rs1050115	*UBXN4*	2	A/G	S	--	T,NA,NA,T	0.17	0.21	0.17	0.15	-0.80 (0.14)	5.69E-09	0.35	0.14	-0.61 (0.24)	**0.01**	0.19
rs9283633	*SI*	3	T/C	**NS**	T>A	T,T,B,T	0.58	0.63	0.46	0.61	0.52 (0.10)	2.03E-07	0.33	0.48	0.29 (0.17)	0.09	0.17
rs185053832	*MGAM*	7	C/A	**NS**	P>T	**D,D,D,D**	0.006	0.01	0.001	0.01	-3.30 (0.49)	1.70E-11	0.63	0.002[7]	-1.26 (3.22)	0.69	0.01
rs17683011	*SLC5A1*	22	A/G	**NS**	N>S	T,T,B,T	0.04	0.06	0.003	0.07	-0.96 (0.19)	3.36E-07	0.31	0.02	-0.94 (0.66)	0.15	0.01
rs17683448	*SLC5A1*	22	C/T	S	--	T,NA,NA,T	0.04	0.06	0.003	0.06	-1.14 (0.21)	5.26E-08	0.38	0.01	-0.97 (0.71)	0.17	0.02

[1]Bonferroni corrected significance threshold = 4.1×10^{-7} (0.05/121,052 SNPs).
[2]A2 is effect allele
[3]NS=nonsynonymous, S=synonymous, I=intron
[4]GERP, SIFT and CADD prediction: D=damaging, T=tolerated otherwise, Polyphen2 prediction: D=probably damaging, P=possibly damaging, B=benign

[5]TGP=1000 genomes allele frequency for Eur (EA) and Afr (AA), AF=allele frequency
[6]*SLC5A10* and *FAM83G* are overlapping genes. These variants are missense variants in *FAM83G* and intronic to *SLC5A10.*
[7]Variants have minor allele count<10

Table 5.2. Significant[1] 1,5-AG (μg/mL) gene-based results in European ancestry sample, validated in African ancestry sample.

		European ancestry (N=6,589)			**African ancestry (N=2,309)**		
Chr	Gene	P-value	cMAF[2]	N SNPs	P-value	cMAF	N SNPs
17	*SLC5A10*	5.13E-64	0.04	58	0.006	0.29	28
17	*FAM83G*	6.24E-17	0.06	56	0.39	0.34	36
7	*MGAM*	8.20E-07	0.09	148	0.06	0.95	98
22	*SLC5A1*	1.10E-06	0.23	48	0.21	0.07	15

[1]Bonferroni corrected significance threshold = 4.0×10^{-6} (0.05/12,504 genes).
[2]cMAF=cumulative minor allele frequency

Figure 5.1. Regional association plots for top hits on chromosome 17 in European ancestry sample.

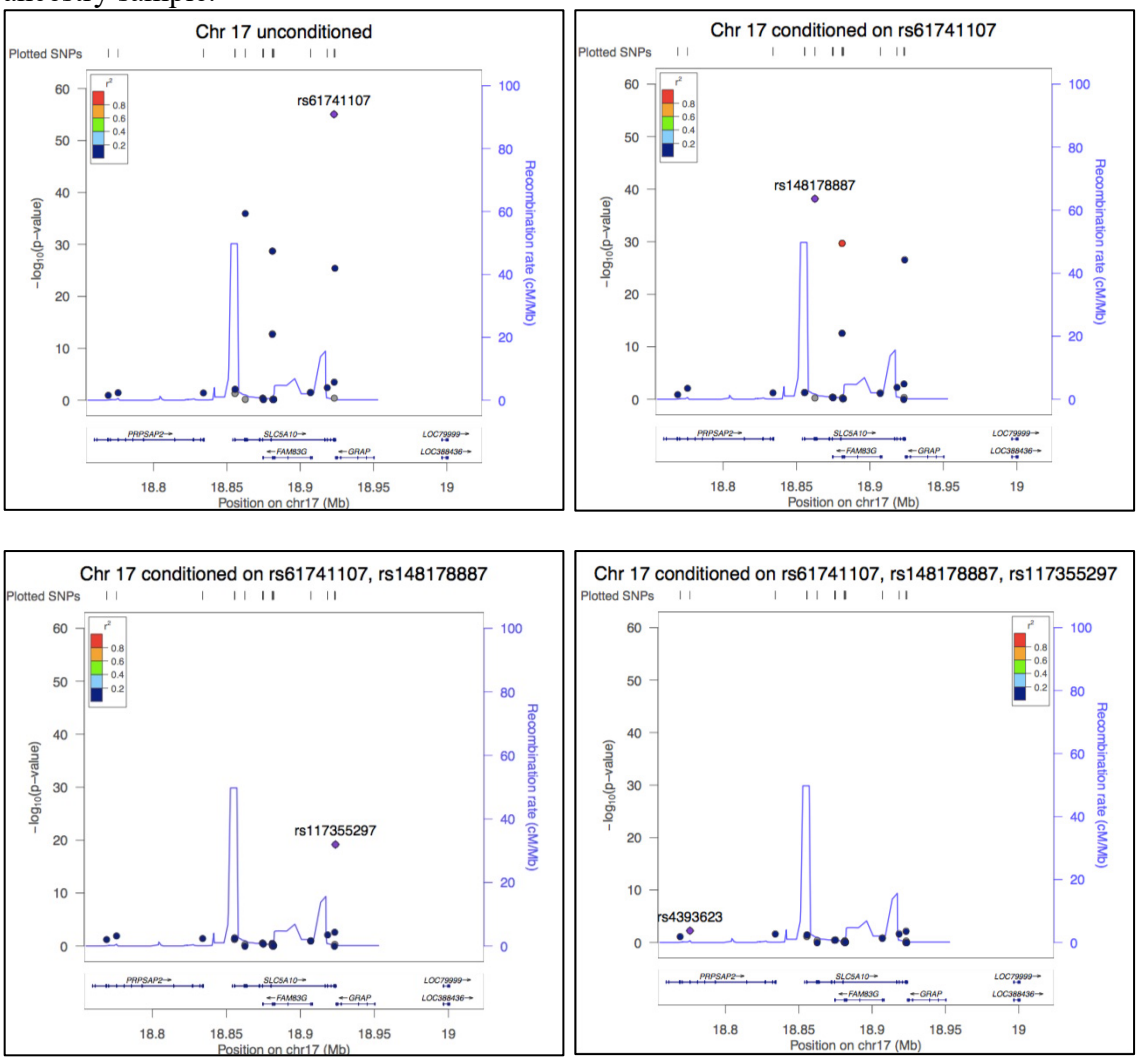
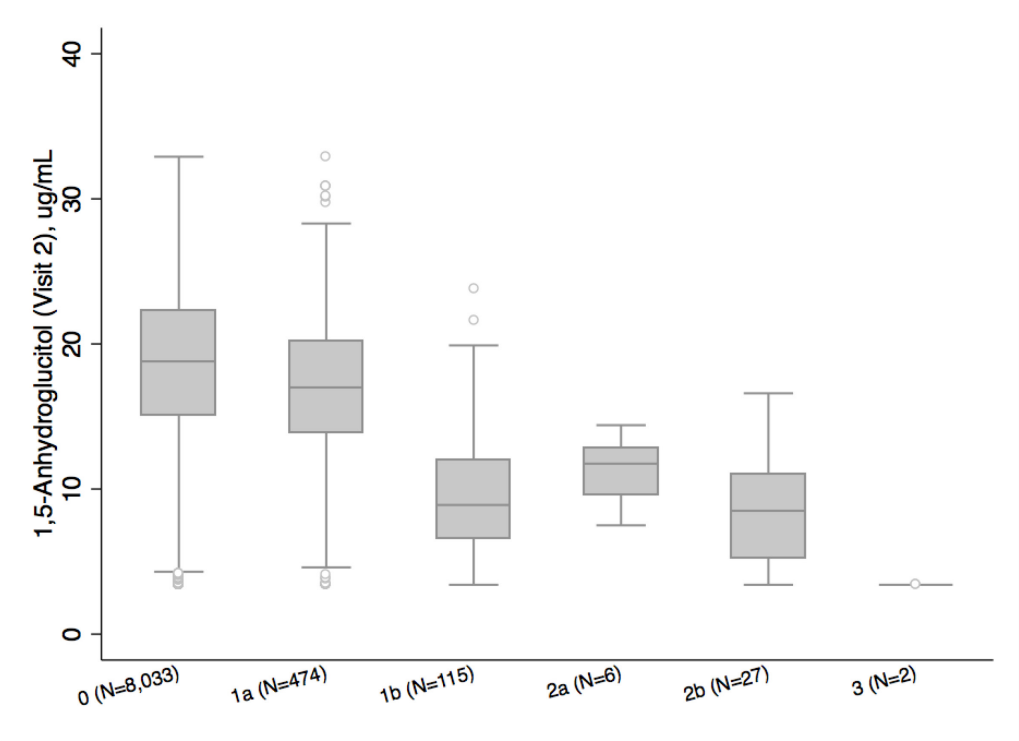


Figure 5.2. Distribution of 1,5-AG by chromosome 17 variants.[1,2]



[1]1,5-AG is winsorized at 1% and 99%
[2]0: no copies of rs61741107, rs148178887 or rs117355297 minor alleles
1a: 1 copy of rs117355297 minor allele
1b: 1 copy of rs61741107 or rs148178887 minor alleles
2a: 2 copies of rs117355297 minor alleles
2b: 1 copy of rs61741107 or rs148178887 + 1 copy of rs117355297 minor allele
3: 1 copy of rs61741107 or rs148178887 + 2 copies of rs117355297 minor allele

Supplemental Table 5.1. Study population characteristics. [1]

	European ancestry (N=6,589)	**African ancestry (N=2,309)**	**Overall (N=8,898)**
Female	55%	63%	57%
Age	57 (5.6)	56 (5.7)	56.7 (5.7)
Fructosamine (µmol/L)	227 (23)	238 (32)	230 (26)
Glycated albumin (%)	12.6 (1.6)	13.6 (2.4)	12.9 (1.9)
1,5-AG (µg/mL)	18.9 (5.8)	17.5 (5.8)	18.5 (5.8)
HbA1c (%)	5.4 (0.52)	5.8 (0.86)	5.5 (0.64)
Fasting glucose (mg/dL)	104 (17)	109 (26)	105 (20)
ARIC study center			
Jackson, Mississippi	0	90%	23%
Forsyth Co, North Carolina	25%	10%	21%
Washington Co, Maryland	30%	0	23%
Minneapolis suburbs, Minneapolis	45%	0	33%

[1]Continuous variables shown as mean (SD) and categorical variables shown as (%)

Supplemental Table 5.2. Association between chromosome 17 SNPs and diabetes status[1]

	European ancestry (N=6,998)				**African ancestry (N=2,704)**			
	Diagnosed diabetes (N=760 cases)		**Diagnosed and undiagnosed diabetes (N=351 cases)**		**Diagnosed diabetes (N=644 cases)**		**Diagnosed and undiagnosed diabetes (N=249 cases)**	
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
rs61741107[2]	0.85 (0.34, 2.12)	0.73	0.95 (0.49, 1.85)	0.88	--	--	--	--
rs148178887	0.34 (0.05, 2.50)	0.29	1.41 (0.63, 3.19)	0.41	1.61 (0.18, 14.55)	0.67	0.85 (0.09, 7.71)	0.89
rs117355297	1.01 (0.69, 1.47)	0.98	1.03 (0.77, 1.34)	0.85	0.96 (0.33, 2.81)	0.95	1.06 (0.45, 2.51)	0.90

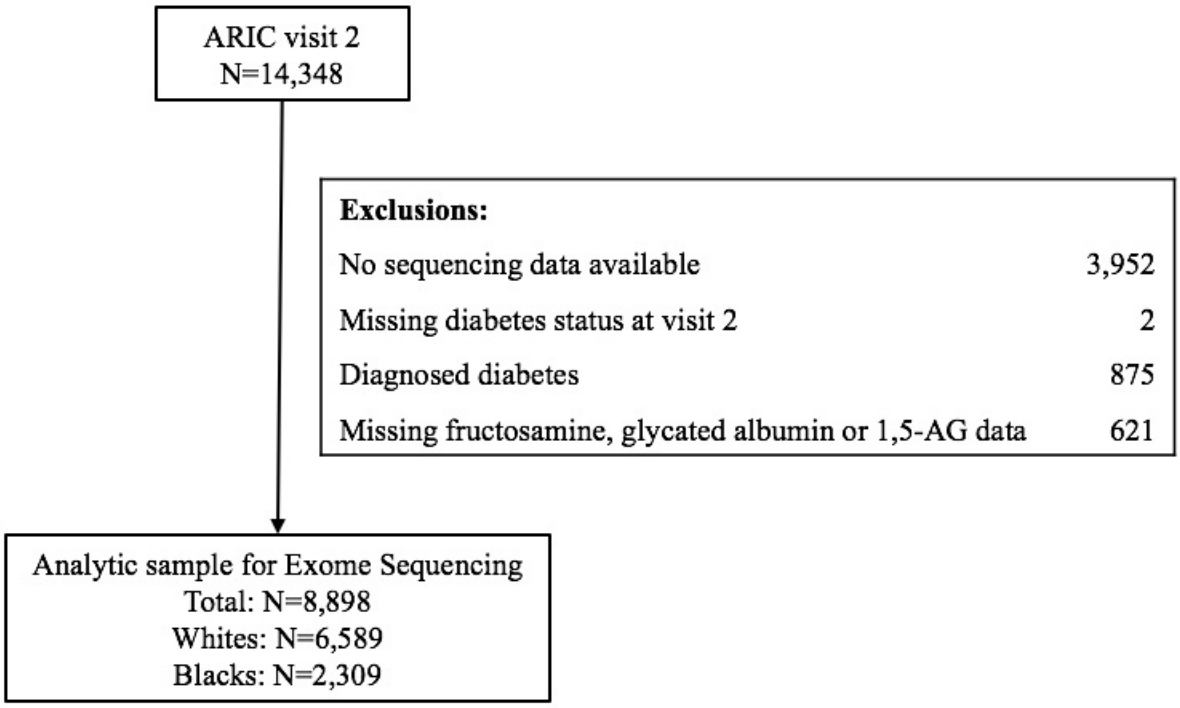
[1]Diagnosed diabetes is defined as self-reported physician diagnosis or use of diabetes medications. Undiagnosed diabetes is defined as fasting glucose ≥126 mg/dL if fasting for ≥8 hours or fasting glucose >200 if not fasting for ≥8 hours).
[2]Only two African ancestry individuals had rs61741107 alleles and thus was too colinear with diabetes to produce a regression estimate.

Supplemental Table 5.3. Chromosome 2 significant results in European ancestry individuals, unconditioned and conditioned on top nonsynonymous variants.

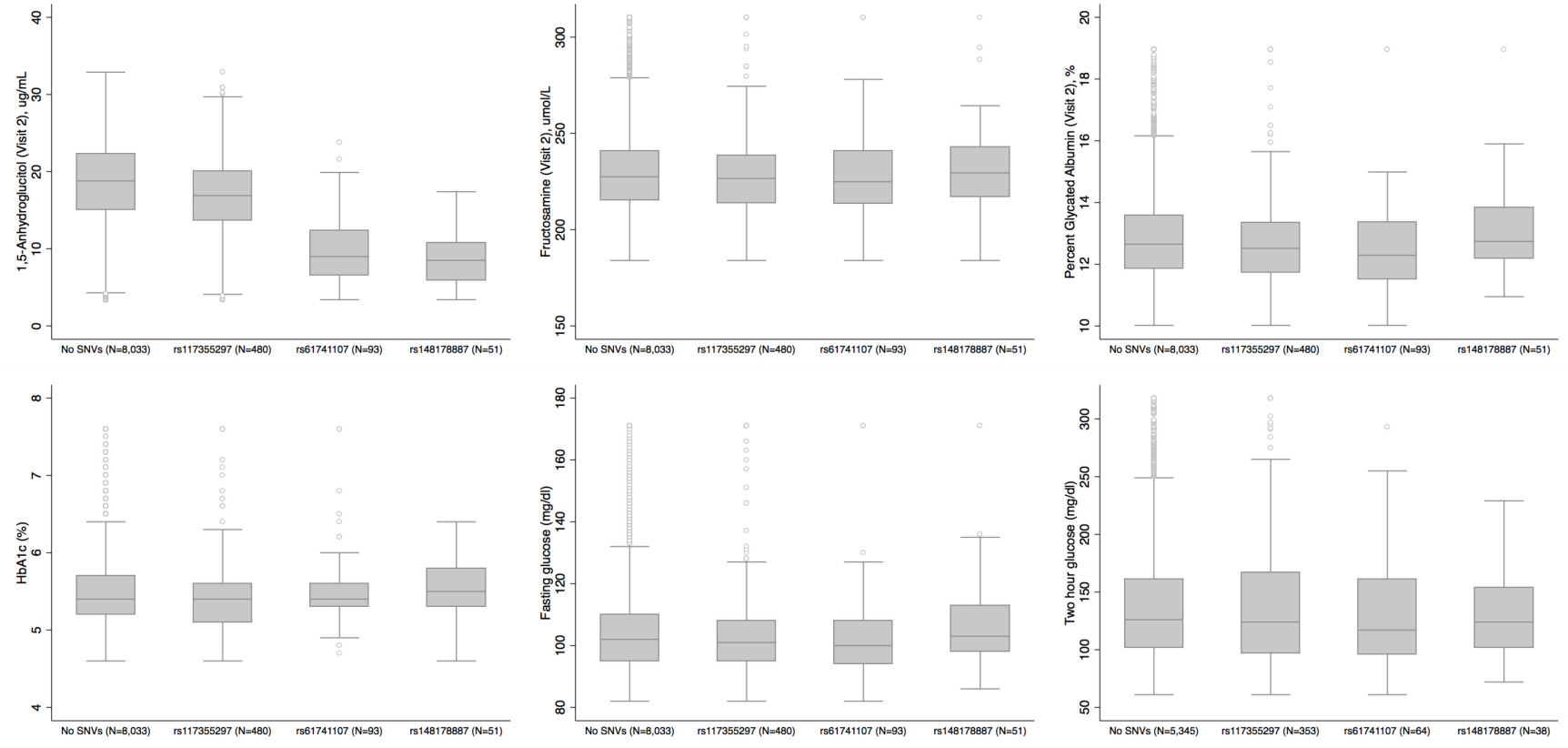
					Unconditioned		**Conditioned on rs961360**		**Conditioned on rs961360 and rs2305165**	
SNP	Gene	Position	A1/A2[2]	Effect AF	Beta (SE)	P-value[1]	Beta (SE)	P-value[1]	Beta (SE)	P-value[1]
rs961360	*R3HDM1*	136393658	A/G	0.15	-0.80 (0.14)	**7.82E-09**	--	--	--	--
rs2305165	*R3HDM1*	136409574	A/C	0.08	-0.85 (0.18)	1.86E-06	-0.98 (0.18)	**5.89E-08**	--	--
rs2304371	*LCT*	136561557	G/A	0.83	0.89 (0.13)	**6.74E-12**	0.79 (0.20)	1.2E-04	0.58 (0.21)	0.006
rs3739022	*LCT*	136562472	G/A	0.10	-1.07 (0.17)	**1.23E-10**	-1.02 (0.17)	**8.99E-10**	-0.78 (0.22)	4.4E-04
rs1050115	*UBXN4*	136511817	A/G	0.15	-0.80 (0.14)	**5.69E-09**	-0.53 (0.27)	0.05	-0.14 (0.28)	0.61
rs10445686	*RAB3GAP1*	135893372	A/G	0.13	-0.79 (0.14)	**3.59E-08**	-0.35 (0.24)	0.14	-0.23 (0.24)	0.33

[1]Bold indicates exome-wide significance (Bonferroni corrected significance threshold = 4.1×10^{-7} (0.05/121,052 SNPs)).

Supplemental Figure 5.1. Sample exclusions

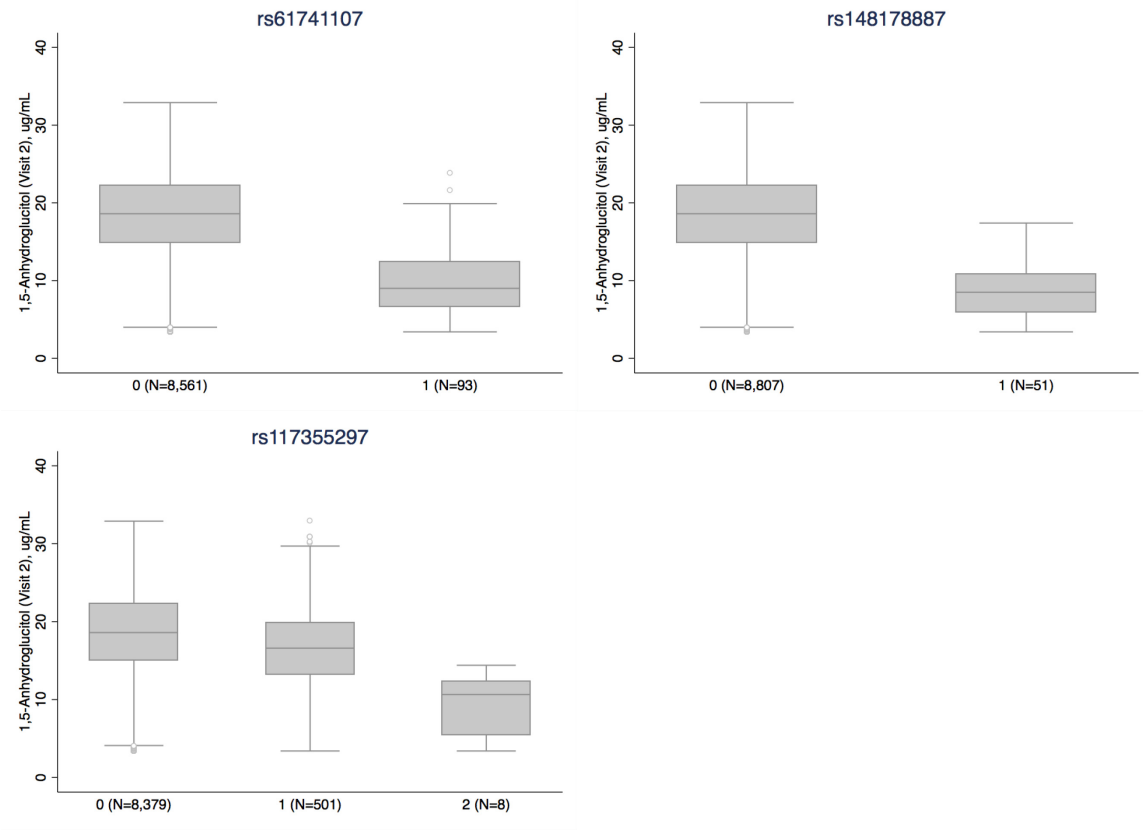


Supplemental Figure 5.2. Distribution of biomarkers by chromosome 17 variants.[1,2]



[1]Biomarkers are winsorized at 1% and 99%.
[2]Two hour glucose was measured at visit 4, other biomarkers were measured at visit 2.

Supplemental Figure 5.3. Distribution of 1,5-AG by chromosome 17 variant genotypes.



Chapter 6: Conclusions

The goal of this work was to better understand the genetics of nontraditional glycemic biomarkers. We hypothesized that genetics would play a substantial role in fructosamine, glycated albumin and 1,5-AG and that both common and rare variants would be associated with these biomarkers. The genetic variants could be diabetes-related, adding insight to disease processes, or they may also be biomarker-specific, representing potential limitations of the biomarkers' ability to accurately reflect blood glucose levels.

6.1 Summary of Key Findings

Heritability

Heritability analyses showed that fructosamine, glycated albumin and 1,5-AG each have genetic variation. The proportion of variance in each biomarker (h^2=0.44-0.55) due to genetics was similar to that of HbA1c (h^2=0.34). In addition, SNP-based heritability showed that common variants make up some of the heritability of these biomarkers but not all, leaving a likely role for rare variants.

Fructosamine and Glycated Albumin Genome Wide Association

GWAS analysis of fructosamine and glycated albumin identified a known diabetes variant in *GCKR*, significantly associated with 1.1% lower levels of percent glycated albumin per minor allele ($p=5.3x10^{-9}$, variance explained=0.3%), and a likely nonglycemic variant in *RCN3,* significantly associated with 1.8% lower fructosamine per

minor allele ($p=5.3x10^{-9}$, variance explained=0.6%). In candidate SNP analysis, few established fasting glucose or HbA1c SNPs were associated with fructosamine or glycated albumin. This work established a role for common variants impacting fructosamine and glycated albumin levels, and showed that these variants represent both glycemic and nonglycemic influences.

Multivariate Phenotype Analysis

In multivariate phenotype analysis, common variants in the *UGT1A* gene region were associated with a joint fructosamine-glycated albumin phenotype ($p=3.18x10^{-8}$) that were not identified through analyses of either phenotype individually. Single phenotype analysis showed a more significant association with fructosamine ($p=6.04x10^{-6}$) than glycated albumin (p=0.72), implying that fructosamine may be driving the multivariate association. The biology of this gene region is complex; it is involved in both diabetes and serum albumin, but the lack of significance with other glycemic biomarkers indicates the effect was more likely driven through albumin. If this is the case, the *UGT1A* association with fructosamine would represent a nonglycemic genetic influence which could alter the clinical interpretation of fructosamine, but further evidence is required to better understand these findings.

1,5-AG Sequencing

Exome sequencing of 1,5-AG showed strong associations with rare and common variants in *SLC5A10*, a glucose transporter exclusively expressed in the kidney. The effect sizes were large, approximately 10 μg/mL lower 1,5-AG levels per risk allele (1,5-

AG range: 3.4-32.8 μg/mL) and seen in both the European American and African American samples. These rare and common variants explained 6% of the total variance of 1,5-AG. Conditional analysis and functional predictions showed strong evidence for an important impact of these variants on 1,5-AG levels. The lack of association with diabetes and lack of change in other glycemic biomarkers (fructosamine, glycated albumin, fasting glucose, HbA1c) by genotype indicate that the *SLC5A10* variants alter 1,5-AG levels in a nonglycemic manner, which may affect the clinical use of 1,5-AG for diabetes monitoring.

6.2 Strengths and Limitations

Strengths and Limitations

A limitation of this work was the level of power to detect the association between genetic variants and nontraditional glycemic biomarkers. Because genetic studies perform millions of association tests (one for each genetic marker-phenotype combination), it is important to control for multiple comparison testing. This is usually done using a strict Bonferroni threshold (0.05/number of tests). Thus, for the exome sequencing and multivariate phenotype analysis, we only considered variants above this threshold to be exome-wide or genome-wide significant. Increasing sample size is a common way to increase power. While the sample size in European ancestry population was relatively large (N=approx.7,000) allowing for detection of variants in exome sequencing and multivariate phenotype as well as heritability, the smaller sample size in African ancestry population (N=approx. 2,000) prevented heritability analysis and likely impeded

detection of rare coding variants in exome sequencing and common variants in the multivariate phenotype analysis.

Another limitation of this dissertation was the lack of available additional cohorts in which to replicate these findings. Due to the Winner's Curse impacting early GWAS, where significant variants found in one study but not in subsequent studies in other cohorts, replication in an additional cohort is necessary to rule out false positive findings from the discovery cohort.[1] In addition, evaluating genetic associations in multiple populations can indicate if a variant is specific to a particular population or genetic ancestry, or if it generalizable across multiple populations. Replication was difficult for fructosamine, glycated albumin and 1,5-AG because they are not commonly collected in epidemiologic studies. The lack of available replication cohorts impeded our ability to replicate our findings from exome sequencing and multivariate phenotype analyses.

Despite these limitations, a strength of this work was its novelty. Few studies have examined the genetics of fructosamine, glycated albumin and 1,5-AG,[2-4] and this work provided new insight into the genetic underpinnings of these biomarkers from heritability to associations. In addition, the sample size among European Americans was substantial and provided sufficient power to detect rare variants associated with 1,5-AG.

Comparison between variants associated with traditional and nontraditional glycemic biomarkers

One goal of this study was to compare newly discovered genetic variants associated with nontraditional glycemic biomarkers (fructosamine, glycated albumin and 1,5-AG) to those found previously associated with traditional glycemic biomarkers (HbA1c and fasting glucose). While we have done this comparison, it is important to

note that the studies which identified associations with fasting glucose and HbA1c were much larger and hence were better powered than the data we had available to find associations with fructosamine, glycated albumin and 1,5-AG. This meant that we had some confidence that the variants unique to the nontraditional glycemic biomarkers were not associated with fasting glucose and HbA1c. However, we could not rule out associations between nontraditional biomarkers and known traditional biomarker genetic variants that may be identified using larger sample sizes with more power to detect associations.

Glycemic biomarkers in the nondiabetic range

Fructosamine, glycated albumin and 1,5-AG have been proposed as complementary, and in some instances, alternative measures of hyperglycemia that can be used to monitor glycemic control in diabetes patients. In this study, however, these biomarkers were evaluated among individuals without diagnosed diabetes. This was necessary to avoid the impact of treated diabetes on hyperglycemia levels, which would negatively bias associations between genetic variants and levels of glycemic biomarkers. However, this also has the effect of truncating the distribution of glycemic biomarkers so that fewer individuals with overt levels of hyperglycemia (extreme biomarker values) were captured. Individuals with undiagnosed diabetes were retained in an attempt to capture some of the higher end of the distribution, but the number of people in that category was limited and their glucose levels did not reflect the highest levels in the distribution. This is a common problem in other fields that study continuous markers of disease and some, such as blood pressure, simply implement a transformation for blood pressure values for individuals on treatment. However, no such transformation is

commonly used in the diabetes field, and implementing one would be arbitrary, particularly for fructosamine, glycated albumin and 1,5-AG, which have not been extensively studied. The genetic variants which impact glucose levels at the nondiabetic range are not necessarily the same variants which impact glucose levels at the diabetic range. This means that there may be diabetes-relevant variants associated with fructosamine, glycated albumin and 1,5-AG that were not able to be detected by this work, and we cannot be certain that the variants identified in this dissertation impact diabetes.

6.3 Future Directions

1,5-AG functional studies

The epidemiologic analyses in this dissertation provide evidence for putative causality of three variants in *SLC5A10* on 1,5-AG levels, but causality, which is important for confirming drug or therapeutic targets, cannot be demonstrated without functional studies. These are studies utilizing cell lines or animal models where the expression of *SLC5A10* can be altered (for example, by CRISPER-Cas9 inactivation) and compared to cells or animals with unaltered *SLC5A10* expression.

More cohorts with biomarker data

To fully understand the genetics of fructosamine, glycated albumin and 1,5-AG, these biomarkers need to be studied in additional populations. If they could be collected in samples which already have genotyping and sequencing data available, our findings could be evaluated for replication and the samples could be meta-analyzed, increasing the power to detect more associations at varying effect sizes.

Specifically, collecting biomarker data in African ancestry cohorts is important to determine if the genetic architecture is similar or different to that of European ancestry studies. HbA1c is differentially impacted by genetic ancestry,[5] and it is important to determine if fructosamine, glycated albumin and 1,5-AG also are altered differently by genetic ancestry. This is particularly essential because average levels of all glycemic biomarkers differ by race, but the social construct of self-identified race does not necessarily reflect biologic genetic ancestry. Race and genetic ancestry are correlated and thus are often conflated, but while genetic ancestry solely reflects biology, race reflects both biology and environmental (i.e., social and cultural) factors. Thus, differences in biomarker levels by race may be due to differences in environmental risk factors rather than differences in biology. In lieu of race-specific clinical recommendations, Leong and Wheeler have proposed genotype-specific recommendations[6] which would focus on differential allele frequencies and biology rather than risk factors associated with race.

Whole-genome sequencing

Whole-exome sequencing has the ability to capture rare variants in the coding region of the genome but is not designed to explain the noncoding region of the genome. While coding variants can impact the protein product of a gene and thus are potentially deleterious, this is not the only type of genetic variation which affects biology. For example, variants in the noncoding region can affect gene regulation, controlling gene expression. In the future, associations between whole-genome sequencing variations including structural variations and fructosamine, glycated albumin and 1,5-AG should be evaluated.

6.4 Public Health Significance

The ultimate goal of this work was to improve treatment and care for those at risk for or diagnosed with type 2 diabetes. An important basic step toward achieving this goal is to know if a person has diabetes or not, and if they do, how well it is managed. This can be achieved by accurately reflecting hyperglycemia, which is measured by biomarkers. The shortcomings of current clinical biomarkers-- fasting glucose and HbA1c-- have led to increasing interest in alternatives or complimentary measures such as fructosamine, glycated albumin and 1,5-AG. However, to use these nontraditional glycemic biomarkers clinically, it is necessary to understand their glycemic (glucose-related) and nonglycemic (biomarker-limitations) determinants. Genetics are a useful tool to elucidate these factors.

Heritability analysis has shown that these biomarkers are under substantial genetic control, which has strengthened the case for the utility in understanding biomarker genetics to explain the variance in fructosamine, glycated albumin and 1,5-AG.

Arguably the most clinically meaningful findings from this dissertation were the variants in *SLC5A10* associated with 1,5-AG. While these variants are rare, they explain 6% of the variance in 1,5-AG, and the large effect sizes mean that an individual with a risk allele in this gene would have greatly reduced 1,5-AG levels: approximately 10 μg/mL per risk allele, where the range of 1,5-AG in this sample was 3.4-32.8 μg/mL. If 1,5-AG was used to measure glycemic control in such individuals, it would produce false positive results. These could result in unnecessary medication use or higher doses than necessary, leading to hypoglycemia. Our results may also suggest a new potential limitation of 1,5-AG in a clinical setting.

Alternatively, utility of 1,5-AG may differ by genotype, and should only be used in individuals without *SLC5A10* risk alleles. While personalized medicine is often thought of in respect to treatment options, it can also be implemented with diagnostics. Clinical genomics is becoming more prevalent, with large healthcare systems such as Geisinger Health System, Kaiser Permanente, and Veteran's Health Administration all collecting genome-wide data on their patients. Where such data exists, it would require little cost to have a message appear in the electronic medical record indicating if a patient's genotype might impact the ability of 1,5-AG to reflect his or her glycemic peaks when a clinician orders a 1,5-AG test. In such a setting, persons with nonglucose related 1,5-AG genotypes may be flagged, and clinicians could tailor their choice of glucose tests accordingly. This is an important piece of the broader precision medicine initiative that can improve treatment by using existing data.

The specific genetic variants associated with fructosamine and glycated albumin have been more difficult to uncover. While the findings from this dissertation are not sufficient to impact the clinical use of fructosamine and glycated albumin, our results highlighted some potential limitations, and are a good first step towards additional understanding of how genetics may impact the clinical interpretation of these biomarkers.

This dissertation has evaluated the role of genetics in nontraditional biomarkers of hyperglycemia – fructosamine, glycated albumin and 1,5-AG – contributing a piece to the large puzzle of improving health for those at risk for and diagnosed with type 2 diabetes.

References

Chapter 1 References

1. Selvin E, Parrinello CM, Sacks DB, Coresh J. Trends in prevalence and control of diabetes in the united states, 1988-1994 and 1999-2010. Ann Intern Med. 2014;160(8):517-525. Accessed Mar 14, 2018. doi: 10.7326/M13-2411.

2. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: A pooled analysis of 751 population-based studies with 4.4 million participants. Lancet. 2016;387(10027):1513-1530. Accessed Mar 14, 2018. doi: 10.1016/S0140-6736(16)00618-8.

3. Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract. 2011;94(3):311-321. Accessed Mar 14, 2018. doi: 10.1016/j.diabres.2011.10.029.

4. Marathe PH, Gao HX, Close KL. American diabetes association standards of medical care in diabetes 2017. J Diabetes. 2017;9(4):320-324. Accessed Jan 22, 2018. doi: 10.1111/1753-0407.12524.

5. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. Lancet. 2017;389(10085):2239-2251. Accessed Mar 14, 2018. doi: 10.1016/S0140-6736(17)30058-2.

6. Rao Kondapally Seshasai S, Kaptoge S, Thompson A, et al. Diabetes mellitus, fasting glucose, and risk of cause-specific death. N Engl J Med. 2011;364(9):829-841. Accessed Mar 14, 2018. doi: 10.1056/NEJMoa1008862.

7. Sarwar N, Gao P, Seshasai SRK, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102

prospective studies. Lancet. 2010;375(9733):2215-2222. Accessed Mar 14, 2018. doi: 10.1016/S0140-6736(10)60484-9.

8. Willemsen G, Ward KJ, Bell CG, et al. The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: The discordant twin (DISCOTWIN) consortium. Twin Res Hum Genet. 2015;18(6):762-771. Accessed Mar 14, 2018. doi: 10.1017/thg.2015.83.

9. Kaprio J, Tuomilehto J, Koskenvuo M, et al. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in finland. Diabetologia. 1992;35(11):1060-1067. Accessed Mar 14, 2018.

10. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia. 1999;42(2):139-145. Accessed Mar 14, 2018.

11. Almgren P, Lehtovirta M, Isomaa B, et al. Heritability and familiality of type 2 diabetes and related quantitative traits in the botnia study. Diabetologia. 2011;54(11):2811-2819. Accessed Mar 14, 2018. doi: 10.1007/s00125-011-2267-5.

12. Flannick J, Florez JC. Type 2 diabetes: Genetic data sharing to advance complex disease research. Nat Rev Genet. 2016;17(9):535-549. Accessed Mar 14, 2018. doi: 10.1038/nrg.2016.56.

13. Scott RA, Scott LJ, Mägi R, et al. An expanded genome-wide association study of type 2 diabetes in europeans. Diabetes. 2017;66(11):2888-2902. Accessed Mar 14, 2018. doi: 10.2337/db16-1253.

14. Goldstein DE, Little RR, Lorenz RA, et al. Tests of glycemia in diabetes. Diabetes Care. 2004;27(7):1761-1773. Accessed Mar 14, 2018.

15. Sacks DB. A1C versus glucose testing: A comparison. Diabetes Care. 2011;34(2):518-523. Accessed Jan 22, 2018. doi: 10.2337/dc10-1546.

16. Sacks DB, John WG. Interpretation of hemoglobin A1c values. JAMA. 2014;311(22):2271-2272. Accessed Jan 22, 2018. doi: 10.1001/jama.2014.6342.

17. Parrinello CM, Selvin E. Beyond HbA1c and glucose: The role of nontraditional glycemic markers in diabetes diagnosis, prognosis, and management. Curr Diab Rep. 2014;14(11):548. Accessed Jan 22, 2018. doi: 10.1007/s11892-014-0548-3.

18. Armbruster DA. Fructosamine: Structure, analysis, and clinical usefulness. Clin Chem. 1987;33(12):2153-2163. Accessed Jan 22, 2018.

19. Buse JB, Freeman JLR, Edelman SV, Jovanovic L, McGill JB. Serum 1,5-anhydroglucitol (GlycoMark): A short-term glycemic marker. Diabetes Technol Ther. 2003;5(3):355-363. Accessed Jan 22, 2018. doi: 10.1089/152091503765691839.

20. Dungan KM. 1,5-anhydroglucitol (GlycoMark) as a marker of short-term glycemic control and glycemic excursions. Expert Rev Mol Diagn. 2008;8(1):9-19. Accessed Jan 22, 2018. doi: 10.1586/14737159.8.1.9.

21. Yamanouchi T, Akanuma Y. Serum 1,5-anhydroglucitol (1,5 AG): New clinical marker for glycemic control. Diabetes Res Clin Pract. 1994;24 Suppl:261. Accessed Jan 22, 2018.

22. Cohen RM, Sacks DB. Comparing multiple measures of glycemia: How to transition from biomarker to diagnostic test? Clin Chem. 2012;58(12):1615-1617. doi: 10.1373/clinchem.2012.196139 [doi].

23. Jun JE, Hur KY, Lee Y-, et al. Glycated albumin predicts the development of early diabetic nephropathy in patients with type 2 diabetes. Diabetes Metab. 2017. Accessed Mar 14, 2018. doi: 10.1016/j.diabet.2017.08.003.

24. Park SB, Kim SS, Kim IJ, et al. Variability in glycated albumin levels predicts the progression of diabetic nephropathy. J Diabetes Complicat. 2017;31(6):1041-1046. Accessed Mar 14, 2018. doi: 10.1016/j.jdiacomp.2017.01.014.

25. Selvin E, Rawlings AM, Grams M, et al. Fructosamine and glycated albumin for risk stratification and prediction of incident diabetes and microvascular complications: A prospective cohort analysis of the atherosclerosis risk in communities (ARIC) study. Lancet Diabetes Endocrinol. 2014;2(4):279-288. doi: 10.1016/S2213-8587(13)70199-2 [doi].

26. Selvin E, Rawlings A, Lutsey P, et al. Association of 1,5-anhydroglucitol with cardiovascular disease and mortality. Diabetes. 2016;65(1):201-208. doi: 10.2337/db15-0607 [doi].

27. Selvin E, Rawlings AM, Grams M, Klein R, Steffes M, Coresh J. Association of 1,5-anhydroglucitol with diabetes and microvascular conditions. Clin Chem. 2014;60(11):1409-1418. doi: 10.1373/clinchem.2014.229427 [doi].

28. Juraschek SP, Steffes MW, Miller ER, Selvin E. Alternative markers of hyperglycemia and risk of diabetes. Diabetes Care. 2012;35(11):2265-2270. Accessed Mar 13, 2018. doi: 10.2337/dc12-0787.

29. Zaccardi F, Kurl S, Pitocco D, Ronkainen K, Laukkanen JA. Serum fructosamine and risk of cardiovascular and all-cause mortality: A 24-year prospective population-based

study. Nutr Metab Cardiovasc Dis. 2015;25(2):236-241. Accessed Mar 14, 2018. doi: 10.1016/j.numecd.2014.09.007.

30. Malmstrom H, Walldius G, Grill V, Jungner I, Gudbjornsdottir S, Hammar N. Fructosamine is a useful indicator of hyperglycaemia and glucose control in clinical and epidemiological studies--cross-sectional and longitudinal experience from the AMORIS cohort. PLoS One. 2014;9(10):e111463. doi: 10.1371/journal.pone.0111463 [doi].

31. Selvin E, Rawlings AM, Lutsey PL, et al. Fructosamine and glycated albumin and the risk of cardiovascular outcomes and death. Circulation. 2015;132(4):269-277. doi: 10.1161/CIRCULATIONAHA.115.015415 [doi].

32. Nathan DM, McGee P, Steffes MW, Lachin JM. Relationship of glycated albumin to blood glucose and HbA1c values and to retinopathy, nephropathy, and cardiovascular outcomes in the DCCT/EDIC study. Diabetes. 2014;63(1):282-290. Accessed Mar 14, 2018. doi: 10.2337/db13-0782.

33. Araki T, Ishikawa Y, Okazaki H, et al. Introduction of glycated albumin measurement for all blood donors and the prevalence of a high glycated albumin level in japan. J Diabetes Investig. 2012;3(6):492-497. Accessed Mar 14, 2018. doi: 10.1111/j.2040-1124.2012.00224.x.

34. Welsh KJ, Kirkman MS, Sacks DB. Role of glycated proteins in the diagnosis and management of diabetes: Research gaps and future directions. Diabetes Care. 2016;39(8):1299-1306. doi: 10.2337/dc15-2727 [doi].

35. Schleicher ED, Olgemöller B, Wiedenmann E, Gerbitz KD. Specific glycation of albumin depends on its half-life. Clin Chem. 1993;39(4):625-628. Accessed Mar 14, 2018.

36. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44(9):991-1005. doi: 10.1038/ng.2385 [doi].

37. Wheeler E, Leong A, Liu C, et al. Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. 2017;14(9):e1002383. Accessed Jan 22, 2018. doi: 10.1371/journal.pmed.1002383.

38. Leong A, Wheeler E. Genetics of HbA1c: A case study in clinical translation. Curr Opin Genet Dev. 2018;50:79-85. Accessed Mar 12, 2018. doi: 10.1016/j.gde.2018.02.008.

39. Chen P, Takeuchi F, Lee J, et al. Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in east asians. Diabetes. 2014;63(7):2551-2562. Accessed Mar 14, 2018. doi: 10.2337/db13-1815.

40. Soranzo N, Sanna S, Wheeler E, et al. Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic pathways. Diabetes. 2010;59(12):3229-3239. Accessed Mar 14, 2018. doi: 10.2337/db10-0502.

41. Shin S, Fauman EB, Petersen A, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46(6):543-550. Accessed Jan 22, 2018. doi: 10.1038/ng.2982.

42. Long T, Hicks M, Yu H, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet. 2017;49(4):568-578. Accessed Jan 22, 2018. doi: 10.1038/ng.3809.

43. Li M, Maruthur NM, Loomis SJ, et al. Genome-wide association study of 1,5-anhydroglucitol identifies novel genetic loci linked to glucose metabolism. Sci Rep. 2017;7(1):2812. Accessed Jan 22, 2018. doi: 10.1038/s41598-017-02287-x.

Chapter 2 References

1. Selvin E, Parrinello CM, Sacks DB, Coresh J. Trends in prevalence and control of diabetes in the united states, 1988-1994 and 1999-2010. Ann Intern Med. 2014;160(8):517-525. Accessed Jan 22, 2018. doi: 10.7326/M13-2411.

2. Marathe PH, Gao HX, Close KL. American diabetes association standards of medical care in diabetes 2017. J Diabetes. 2017;9(4):320-324. Accessed Jan 22, 2018. doi: 10.1111/1753-0407.12524.

3. Parrinello CM, Selvin E. Beyond HbA1c and glucose: The role of nontraditional glycemic markers in diabetes diagnosis, prognosis, and management. Curr Diab Rep. 2014;14(11):548. Accessed Jan 22, 2018. doi: 10.1007/s11892-014-0548-3.

4. Sacks DB. A1C versus glucose testing: A comparison. Diabetes Care. 2011;34(2):518-523. Accessed Jan 22, 2018. doi: 10.2337/dc10-1546.

5. Sacks DB, John WG. Interpretation of hemoglobin A1c values. JAMA. 2014;311(22):2271-2272. Accessed Jan 22, 2018. doi: 10.1001/jama.2014.6342.

6. Goldstein DE, Little RR, Lorenz RA, et al. Tests of glycemia in diabetes. Diabetes Care. 2004;27(7):1761-1773. Accessed Jan 22, 2018.

7. Armbruster DA. Fructosamine: Structure, analysis, and clinical usefulness. Clin Chem. 1987;33(12):2153-2163. Accessed Jan 22, 2018.

8. Buse JB, Freeman JLR, Edelman SV, Jovanovic L, McGill JB. Serum 1,5-anhydroglucitol (GlycoMark): A short-term glycemic marker. Diabetes Technol Ther. 2003;5(3):355-363. Accessed Jan 22, 2018. doi: 10.1089/152091503765691839.

9. Dungan KM. 1,5-anhydroglucitol (GlycoMark) as a marker of short-term glycemic control and glycemic excursions. Expert Rev Mol Diagn. 2008;8(1):9-19. Accessed Jan 22, 2018. doi: 10.1586/14737159.8.1.9.

10. Yamanouchi T, Akanuma Y. Serum 1,5-anhydroglucitol (1,5 AG): New clinical marker for glycemic control. Diabetes Res Clin Pract. 1994;24 Suppl:261. Accessed Jan 22, 2018.

11. Shin S, Fauman EB, Petersen A, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46(6):543-550. Accessed Jan 22, 2018. doi: 10.1038/ng.2982.

12. Meigs JB, Panhuysen CIM, Myers RH, Wilson PWF, Cupples LA. A genome-wide scan for loci linked to plasma levels of glucose and HbA(1c) in a community-based sample of caucasian pedigrees: The framingham offspring study. Diabetes. 2002;51(3):833-840. Accessed Jan 22, 2018.

13. Pilia G, Chen W, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 sardinians. PLoS Genet. 2006;2(8):e132. Accessed Jan 22, 2018. doi: 10.1371/journal.pgen.0020132.

14. Watanabe RM, Valle T, Hauser ER, et al. Familiality of quantitative metabolic traits in finnish families with non-insulin-dependent diabetes mellitus. finland-united states investigation of NIDDM genetics (FUSION) study investigators. Hum Hered. 1999;49(3):159-168. Accessed Jan 22, 2018. doi: 10.1159/000022865.

15. Mills GW, Avery PJ, McCarthy MI, et al. Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased

susceptibility to type 2 diabetes. Diabetologia. 2004;47(4):732-738. Accessed Jan 22, 2018. doi: 10.1007/s00125-004-1338-2.

16. Hsueh WC, Mitchell BD, Aburomia R, et al. Diabetes in the old order amish: Characterization and heritability analysis of the amish family diabetes study. Diabetes Care. 2000;23(5):595-601. Accessed Jan 22, 2018.

17. Simonis-Bik AMC, Eekhoff EMW, Diamant M, et al. The heritability of HbA1c and fasting blood glucose in different measurement settings. Twin Res Hum Genet. 2008;11(6):597-602. Accessed Jan 22, 2018. doi: 10.1375/twin.11.6.597.

18. Snieder H, Sawtell PA, Ross L, Walker J, Spector TD, Leslie RD. HbA(1c) levels are genetically determined even in type 1 diabetes: Evidence from healthy and diabetic twins. Diabetes. 2001;50(12):2858-2863. Accessed Jan 22, 2018.

19. Mitchell BD, Kammerer CM, Blangero J, et al. Genetic and environmental contributions to cardiovascular risk factors in mexican americans. the san antonio family heart study. Circulation. 1996;94(9):2159-2170. Accessed Jan 22, 2018.

20. Long T, Hicks M, Yu H, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet. 2017;49(4):568-578. Accessed Jan 22, 2018. doi: 10.1038/ng.3809.

21. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. Am J Hum Genet. 2012;91(6):1011-1021. Accessed Jan 22, 2018. doi: 10.1016/j.ajhg.2012.10.010.

22. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7):565-569. Accessed Jan 22, 2018. doi: 10.1038/ng.608.

23. The atherosclerosis risk in communities (ARIC) study: Design and objectives. the ARIC investigators. Am J Epidemiol. 1989;129(4):687-702. Accessed Jan 22, 2018.

24. Selvin E, Coresh J, Zhu H, Folsom A, Steffes MW. Measurement of HbA1c from stored whole blood samples in the atherosclerosis risk in communities study. J Diabetes. 2010;2(2):118-124. Accessed Jan 22, 2018. doi: 10.1111/j.1753-0407.2010.00070.x.

25. Purcell S, Neale B, Tood-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. AJHG. 2007 Sep;81(3):559-75. Epub 2007 Jul 25.

26. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74. Accessed Jan 22, 2018. doi: 10.1038/nature15393.

27. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998;62(5):1198-1211. Accessed Jan 22, 2018. doi: 10.1086/301844.

28. Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ. Reevaluation of SNP heritability in complex human traits. Nat Genet. 2017;49(7):986-992. Accessed Jan 22, 2018. doi: 10.1038/ng.3865.

29. McAdams DeMarco M, Coresh J, Woodward M, et al. Hypertension status, treatment, and control among spousal pairs in a middle-aged adult cohort. Am J Epidemiol. 2011 Oct 1;174(7):790-6. doi: 10.1093/aje/kwr167. Epub 2011 Aug 12.

30. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. Nat Genet. 2017;49(9):1304-1310. Accessed Jan 22, 2018. doi: 10.1038/ng.3941.

31. Yang J, Bakshi A, Zhu Z, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47(10):1114-1120. Accessed Jan 22, 2018. doi: 10.1038/ng.3390.

32. Lee SH, Yang J, Chen G, et al. Estimation of SNP heritability from dense genotype data. Am J Hum Genet. 2013;93(6):1151-1155. Accessed Jan 22, 2018. doi: 10.1016/j.ajhg.2013.10.015.

33. Wheeler E, Leong A, Liu C, et al. Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. 2017;14(9):e1002383. Accessed Jan 22, 2018. doi: 10.1371/journal.pmed.1002383.

34. Li M, Maruthur NM, Loomis SJ, et al. Genome-wide association study of 1,5-anhydroglucitol identifies novel genetic loci linked to glucose metabolism. Sci Rep. 2017;7(1):2812. Accessed Jan 22, 2018. doi: 10.1038/s41598-017-02287-x.

35. Loomis SJ, Li M, et al. Genome-wide association study of serum fructosamine and glycated albumin in adults without diagnosed diabetes: results from the Atherosclerosis Risk in Communities Study. Under review at Diabetes.

36. Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. PLoS Genet. 2012;8(3):e1002637. doi: 10.1371/journal.pgen.1002637. Epub 2012 Mar 29.

37. Anguizola J, Matsuda R, Barnaby OS, et al. Review: Glycation of human serum albumin. Clin Chim Acta. 2013;425:64-76. Accessed Jan 22, 2018. doi: 10.1016/j.cca.2013.07.013.

38. Cohen MP. Clinical, pathophysiological and structure/function consequences of modification of albumin by amadori-glucose adducts. Biochim Biophys Acta. 2013;1830(12):5480-5485. Accessed Jan 22, 2018. doi: 10.1016/j.bbagen.2013.04.024.

Chapter 3 References

1. Marathe PH, Gao HX, Close KL. American Diabetes Association Standards of Medical Care in Diabetes 2017. J Diabetes. 2017 Jan;40 supplem:1–142.
2. Sacks DB. A1C versus glucose testing: A comparison. Diabetes Care. 2011;34(2):518–23.
3. Selvin E, Rawlings AM, Grams M, Klein R, Sharrett a. R, Steffes M, et al. Fructosamine and glycated albumin for risk stratification and prediction of incident diabetes and microvascular complications: A prospective cohort analysis of the Atherosclerosis Risk in Communities (ARIC) study. Lancet Diabetes Endocrinol. 2014;2(4):279–88.
4. Selvin E, Rawlings AM, Lutsey PL, Maruthur N, Pankow JS, Steffes M, et al. Fructosamine and glycated albumin and the risk of cardiovascular outcomes and death. Circulation. 2015;132(4):269–77.
5. Johns Hopkins POC-IT Guides: alternative markers of glycemia: fructosamine, glycated albumin, 1,5-AG. Available at: http://www. hopkinsguides.com/hopkins/ub/view/Johns_Hopkins_Diabetes_ Guide/547055/all/Alternative_markers_of_glycemia:_fructosamine__ glycated_albumin__1_5_AG. Accessed 20 Jan 2018
6. Joslin Diabetes Center: Home blood glucose (sugar) monitoring, hemoglobin A1C testing, and fructosamine tests. Available at: http://www.joslin.org/info/home_blood_glucose_sugar_ monitoring_hemoglobin_a1c_testing_and_fructosamine_tests. html. Accessed 20

Jan 2018.

7. He X, Ying L, Ma X, Shen Y, Su H, Peng J, et al. An additional measurement of glycated albumin can help prevent missed diagnosis of diabetes in Chinese population. Clin Chim Acta. 2017 Dec;475:188-192.
8. Malmström H, Walldius G, Grill V, Jungner I, Gudbjörnsdottir S, Hammar N. Fructosamine is a useful indicator of hyperglycaemia and glucose control in clinical and epidemiological studies--cross-sectional and longitudinal experience from the AMORIS cohort. PLoS One. 2014 Oct 29;9(10):e111463.
9. Welsh KJ, Kirkman MS, Sacks DB. Role of Glycated Proteins in the Diagnosis and Management of Diabetes: Research Gaps and Future Directions. Diabetes Care. 2016 Aug;39(8):1299-306.
10. Danese E, Montagnana M, Nouvenne A, Lippi G. Advantages and pitfalls of fructosamine and glycated albumin in the diagnosis and treatment of diabetes. J Diabetes Sci Technol. 2015 Mar;9(2):169-76.
11. Koga M, Kasayama S. Clinical impact of glycated albumin as another glycemic control marker. Endocr J. 2010;57(9):751-62.
12. Rubinow KB, Hirsch IB. Reexamining metrics for glucose control. JAMA. 2011 Mar 16;305(11):1132-3.
13. Cohen RM, Sacks DB. Comparing multiple measures of glycemia: how to transition from biomarker to diagnostic test? Clin Chem. 2012 Dec;58(12):1615-7.
14. Parrinello, C. M. & Selvin, E. Beyond HbA1c and Glucose: the Role of Nontraditional Glycemic Markers in Diabetes Diagnosis, Prognosis, and Management. *Curr. Diab. Rep.* 14, (2014).

15. Armbruster DA. Fructosamine: structure, analysis, and clinical usefulness. Clin Chem. 1987;33(12):2153–63.

16. Sacks, D. B. A1C versus glucose testing: A comparison. *Diabetes Care* 34**,** 518–523 (2011).

17. Araki T, Ishikawa Y, Okazaki H, Tani Y, Toyooka S, Satake M, et al. Introduction of glycated albumin measurement for all blood donors and the prevalence of a high glycated albumin level in Japan. J Diabetes Investig. 2012;3(6):492–7.

18. Burdett T (EBI), Hall PN (NHGRI), Hastings E (EBI), Hindorff LA (NHGRI), Junkins HA (NHGRI), Klemm AK (NHGRI), MacArthur J (EBI), Manolio TA (NHGRI), Morales J (EBI) PH (EBI) and WD (EBI). The NHGRI-EBI Catalog of published genome-wide association studies. [Internet]. [cited 2017 Jun 30]. Available from: www.ebi.ac.uk/gwas

19. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common Variants at 10 Genomic Loci Influence Hemoglobin A1C Levels via Glycemic and Nonglycemic Pathways. Diabetes. 2010;59(12):3229–39.

20. Chen P, Takeuchi F, Lee JY, Li H, Wu JY, Liang J, et al. Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. Diabetes. 2014;63(7):2551–62.

21. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. 2017 Sep 12;14(9):e1002383.

22. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al.

New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010 Feb;42(2):105–16.

23. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012 Sep;44(9):991–1005.

24. Liu CT, Raghavan S, Maruthur N, Kabagambe EK, Hong J, Ng MC, et al.Trans-ethnic Meta-analysis and Functional Annotation Illuminates the Genetic Architecture of Fasting Glucose and Insulin. Am J Hum Genet. 2016 Jul 7;99(1):56-75

25. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol. 1989 Apr;129(4):687–702.

26. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6).

27. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–9.

28. Selvin E, Coresh J, Zhu H, Folsom A, Steffes MW. Measurement of HbA1c from stored whole blood samples in the Atherosclerosis Risk in Communities study. J Diabetes. 2010;2(2):118–24.

29. Marchini J, Howie BN, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat

Genet. 2007;39(7):906–13.

30. Rosner B. Fundamentals of Biostatistics. Thomson-Brooks/Cole, Duxbury; 2006.

31. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun. 2016;7:10023.

32. Magi R, Morris AP: GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 2010, 11:288.

33. Bulik-Sullivan, et al. LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. Nature Genetics, 2015.

34. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs DR, et al. Cardia: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol. 1988;41(11):1105–16.

35. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010;34(8):816–34.

36. Aulchenko YS, Struchalin M V, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics. 2010;11:134.

37. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNiPA: An interactive, genetic variant-centered annotation browser. Bioinformatics. 2015;31(8):1334–6.

38. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Sep 1;26(17):2190-1

39. Scott RA, Scott LJ, Mägi R, Marullo L, Gaulton KJ, Kaakinen M, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes. 2017 Nov;66(11):2888-2902.

40. Bi M, Kao WH, Boerwinkle E, Hoogeveen RC, Rasmussen-Torvik LJ, Astor BC, et al. Association of rs780094 in GCKR with metabolic traits and incident diabetes and cardiovascular disease: the ARIC Study. PLoS One. 2010 Jul 22;5(7):e11690.

41. Wang H, Liu L, Zhao J, Cui G, Chen C, Ding H, Wang DW. Large scale meta-analyses of fasting plasma glucose raising variants in GCK, GCKR, MTNR1B and G6PC2 and their impacts on type 2 diabetes mellitus risk. PLoS One. 2013 Jun 28;8(6):e67665.

42. Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. Hum Mol Genet. 2015;24(July):ddv264.

43. Köttgen A, Pattaro C, Böger CA, Fuchsberger C, Olden M, Glazer NL, et al. New loci associated with kidney function and chronic kidney disease. Nat Genet. 2010 May;42(5):376-84.

44. Sparsø T, Andersen G, Nielsen T, Burgdorf KS, Gjesing AP, Nielsen AL, et al. The GCKR rs780094 polymorphism is associated with elevated fasting serum triacylglycerol, reduced fasting and OGTT-related insulinaemia, and reduced risk of type 2 diabetes. Diabetologia. 2008 Jan;51(1):70-5.

45. Umeno J, Asano K, Matsushita T, Matsumoto T, Kiyohara Y, Iida M, et al. Meta-analysis of published studies identified eight additional common susceptibility loci for Crohn's disease and ulcerative colitis. Inflamm Bowel Dis. 2011 Dec;17(12):2407-15.

46. Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PR, Orho-Melander M, Gloyn AL. The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. Hum Mol Genet. 2009 Nov 1;18(21):4081-8.

47. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nature Genetics. 29 May 2013. 45(6):580-5.

48. Honoré B. The rapidly expanding CREC protein family: members, localization, function, and role in disease. Bioessays. 2009;31(3):262–77.

49. Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, Daigo Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet. Nature Publishing Group; 2010;42(3):210–5.

50. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, et al. Large-scale genome-wide association studies in east Asians identify new genetic loci influencing metabolic traits. Nat Genet. 2011;43(10):990–5.

Chapter 4 References

1. Goldstein DE, Little RR, Lorenz RA, et al. Tests of glycemia in diabetes. Diabetes Care. 2004;27(7):1761-1773. Accessed Mar 14, 2018.

2. Parrinello CM, Selvin E. Beyond HbA1c and glucose: The role of nontraditional glycemic markers in diabetes diagnosis, prognosis, and management. Curr Diab Rep. 2014;14(11):548. Accessed Jan 22, 2018. doi: 10.1007/s11892-014-0548-3.

3. Malmstrom H, Walldius G, Grill V, Jungner I, Gudbjornsdottir S, Hammar N. Fructosamine is a useful indicator of hyperglycaemia and glucose control in clinical and epidemiological studies--cross-sectional and longitudinal experience from the AMORIS cohort. PLoS One. 2014;9(10):e111463. doi: 10.1371/journal.pone.0111463 [doi].

4. Welsh KJ, Kirkman MS, Sacks DB. Role of glycated proteins in the diagnosis and management of diabetes: Research gaps and future directions. Diabetes Care. 2016;39(8):1299-1306. doi: 10.2337/dc15-2727 [doi].

5. Danese E, Montagnana M, Nouvenne A, Lippi G. Advantages and pitfalls of fructosamine and glycated albumin in the diagnosis and treatment of diabetes. J Diabetes Sci Technol. 2015;9(2):169-176. doi: 10.1177/1932296814567227 [doi].

6. Koga M, Kasayama S. Clinical impact of glycated albumin as another glycemic control marker. Endocr J. 2010;57(9):751-762. doi: JST.JSTAGE/endocrj/K10E-138 [pii].

7. Cohen RM, Sacks DB. Comparing multiple measures of glycemia: How to transition from biomarker to diagnostic test? Clin Chem. 2012;58(12):1615-1617. doi: 10.1373/clinchem.2012.196139 [doi].

8. Armbruster DA. Fructosamine: Structure, analysis, and clinical usefulness. Clin Chem. 1987;33(12):2153-2163. Accessed Jan 22, 2018.

9. Wheeler E, Leong A, Liu C, et al. Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. 2017;14(9):e1002383. Accessed Jan 22, 2018. doi: 10.1371/journal.pmed.1002383.

10. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012;44(9):991-1005. doi: 10.1038/ng.2385 [doi].

11. Shen X, Klarić L, Sharapov, S. Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. Nat Commun. 2017 Sep 6;8(1):447. doi: 10.1038/s41467-017-00453-3.

12. Shriner, D. Moving toward system genetics through multiple trait analysis in genome-wide association studies. Front Genet. 2012 Jan 16;3:1. doi: 10.3389/fgene.2012.00001.

13. Ray D, Pankow JS, Basu S. USAT: A unified score-based association test for multiple phenotype-genotype analysis. Genet Epidemiol. 2016;40(1):20-34. doi: 10.1002/gepi.21937 [doi].

14. The atherosclerosis risk in communities (ARIC) study: Design and objectives. the ARIC investigators. Am J Epidemiol. 1989;129(4):687-702. Accessed Jan 22, 2018.

15. Ray D, Boehnke M. Methods for meta-analysis of multiple traits using GWAS summary statistics. Genet Epidemiol. 2018 Mar;42(2):134-145. doi: 10.1002/gepi.22105.

16. Purcell S, Neale B, Tood-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. AJHG. 2007 Sep;81(3):559-75. Epub 2007 Jul 25.

17. Bosma PJ, Chowdhury JR, Bakker C, et al. The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in gilbert's syndrome. N Engl J Med. 1995;333(18):1171-1175. Accessed Mar 15, 2018. doi: 10.1056/NEJM199511023331802.

18. Bielinski SJ, Chai HS, Pathak J, et al. Mayo genome consortia: A genotype-phenotype resource for genome-wide association studies with an application to the analysis of circulating bilirubin levels. Mayo Clin Proc. 2011;86(7):606-614. Accessed Mar 15, 2018. doi: 10.4065/mcp.2011.0178.

19. Johnson AD, Kavousi M, Smith AV, et al. Genome-wide association meta-analysis for total serum bilirubin levels. Hum Mol Genet. 2009;18(14):2700-2710. Accessed Mar 15, 2018. doi: 10.1093/hmg/ddp202.

20. Sanna S, Busonero F, Maschio A, et al. Common variants in the SLCO1B3 locus are associated with bilirubin levels and unconjugated hyperbilirubinemia. Hum Mol Genet. 2009;18(14):2711-2718. Accessed Mar 15, 2018. doi: 10.1093/hmg/ddp203.

21. Cheriyath P, Gorrepati VS, Peters I, et al. High total bilirubin as a protective factor for diabetes mellitus: An analysis of NHANES data from 1999 - 2006. J Clin Med Res. 2010;2(5):201-206. Accessed Mar 15, 2018. doi: 10.4021/jocmr425w.

22. Inoguchi T, Sasaki S, Kobayashi K, Takayanagi R, Yamada T. Relationship between gilbert syndrome and prevalence of vascular complications in patients with diabetes. JAMA. 2007;298(12):1398-1400. Accessed Mar 15, 2018. doi: 10.1001/jama.298.12.1398-b.

23. Ko GT, Chan JC, Woo J, et al. Serum bilirubin and cardiovascular risk factors in a chinese population. J Cardiovasc Risk. 1996;3(5):459-463. Accessed Mar 15, 2018.

24. Vítek L. The role of bilirubin in diabetes, metabolic syndrome, and cardiovascular diseases. Front Pharmacol. 2012;3:55. Accessed Mar 15, 2018. doi: 10.3389/fphar.2012.00055.

Chapter 5 References

1. Buse JB, Freeman JLR, Edelman SV, Jovanovic L, McGill JB. Serum 1,5-anhydroglucitol (GlycoMark): A short-term glycemic marker. *Diabetes Technol Ther*. 2003;5(3):355-363. Accessed Jan 22, 2018. doi: 10.1089/152091503765691839.

2. Dungan KM. 1,5-anhydroglucitol (GlycoMark) as a marker of short-term glycemic control and glycemic excursions. *Expert Rev Mol Diagn*. 2008;8(1):9-19. Accessed Jan 22, 2018. doi: 10.1586/14737159.8.1.9.

3. Selvin E, Rawlings AM, Grams M, Klein R, Steffes M, Coresh J. Association of 1,5-anhydroglucitol with diabetes and microvascular conditions. *Clin Chem*. 2014;60(11):1409-1418. doi: 10.1373/clinchem.2014.229427 [doi].

4. Selvin E, Rawlings A, Lutsey P, et al. Association of 1,5-anhydroglucitol with cardiovascular disease and mortality. *Diabetes*. 2016;65(1):201-208. doi: 10.2337/db15-0607 [doi].

5. Ikeda N, Hara H, Hiroi Y, Nakamura M. Impact of serum 1,5-anhydro-d-glucitol level on prediction of major adverse cardiac and cerebrovascular events in non-diabetic patients without coronary artery disease. *Atherosclerosis*. 2016;253:1-6. doi: S0021-9150(16)31265-5 [pii].

6. Li M, Maruthur NM, Loomis SJ, et al. Genome-wide association study of 1,5-anhydroglucitol identifies novel genetic loci linked to glucose metabolism. *Sci Rep*. 2017;7(1):2812. Accessed Jan 22, 2018. doi: 10.1038/s41598-017-02287-x.

7. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014;46(6):543-550. doi: 10.1038/ng.2982 [doi].

8. Liu CT, Raghavan S, Maruthur N, et al. Trans-ethnic meta-analysis and functional annotation illuminates the genetic architecture of fasting glucose and insulin. *Am J Hum Genet*. 2016;99(1):56-75. doi: 10.1016/j.ajhg.2016.05.006 [doi].

9. Wheeler E, Leong A, Liu C, et al. Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med*. 2017;14(9):e1002383. Accessed Jan 22, 2018. doi: 10.1371/journal.pmed.1002383.

10. The atherosclerosis risk in communities (ARIC) study: Design and objectives. the ARIC investigators. *Am J Epidemiol*. 1989;129(4):687-702. Accessed Jan 22, 2018.

11. Juraschek SP, Steffes MW, Miller ER, Selvin E. Alternative markers of hyperglycemia and risk of diabetes. *Diabetes Care*. 2012;35(11):2265-2270. Accessed Mar 13, 2018. doi: 10.2337/dc12-0787.

12. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. 2010;26(5):589-595. doi: 10.1093/bioinformatics/btp698 [doi].

13. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814.

14. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi: 10.1038/nmeth0410-248 [doi].

15. Cooper GM, Stone EA, Asimenos G, et al. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15(7):901-913. doi: gr.3577405 [pii].

16. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. 2014;46(3):310-315. http://www.ncbi.nlm.nih.gov/pubmed/24487276. doi: 10.1038/ng.2892.

17. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-2337. doi: 10.1093/bioinformatics/btq419 [doi].

18. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi: 10.1038/ng.2653 [doi].

19. Long T, Hicks M, Yu H, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet*. 2017;49(4):568-578. Accessed Jan 22, 2018. doi: 10.1038/ng.3809.

20. Warren B, Lee AK, Ballantyne CM, Hoogeveen RC, Pankow JS, Grams M, Köttgen A, Selvin E. Diagnostic performance of 1,5-anhydroglucitol compared to 2-hour glucose in the Atherosclerosis Risk in Communities Study. Manuscript in preparation.

21. Chen P, Takeuchi F, Lee JY, Li H, Wu JY, Liang J, et al. Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. Diabetes. 2014;63(7):2551–62.

22. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common Variants at 10 Genomic Loci Influence Hemoglobin A1C Levels via Glycemic and Nonglycemic Pathways. Diabetes. 2010;59(12):3229–39.

Chapter 6 References

1. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. Epidemiology. 2008;19(5):658. Accessed Mar 12, 2018. doi: 10.1097/EDE.0b013e318181b865.

2. Li M, Maruthur NM, Loomis SJ, et al. Genome-wide association study of 1,5-anhydroglucitol identifies novel genetic loci linked to glucose metabolism. Sci Rep. 2017;7(1):2812. Accessed Jan 22, 2018. doi: 10.1038/s41598-017-02287-x.

3. Long T, Hicks M, Yu H, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet. 2017;49(4):568-578. Accessed Jan 22, 2018. doi: 10.1038/ng.3809.

4. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46(6):543-550. doi: 10.1038/ng.2982 [doi].

5. Wheeler E, Leong A, Liu C, et al. Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. 2017;14(9):e1002383. Accessed Jan 22, 2018. doi: 10.1371/journal.pmed.1002383.

6. Leong A, Wheeler E. Genetics of HbA1c: A case study in clinical translation. Curr Opin Genet Dev. 2018;50:79-85. Accessed Mar 12, 2018. doi: 10.1016/j.gde.2018.02.008.

Stephanie J. Loomis, MPH

stephloomis@gmail.com
812 N. Calvert St. • Baltimore, MD • 21202 • 860.912.7169

Qualifications Summary:

- 11 years experience in genetic and genetic epidemiology research, including genome-wide association study (GWAS), imputed GWAS, exome array, SNP heritability, exome sequencing analysis, SNP score analysis, gene x environment and metabolic pathway analysis
- Ability to successfully collaborate with interdisciplinary investigators and as part of consortia, resulting in over 30 publications in journals such as Nature and Nature Genetics
- Proficient in biostatistical genetics programs including PLINK, SAS, Stata, R, ProbABEL, LDAK, METAL, and R, as well as the Linux operating system

Education:

PhD candidate, Epidemiology
Genetic Epidemiology concentration — Aug 2014-present
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

MPH, Epidemiology/Biostatistics concentration — May 2008
Tufts University School of Medicine, Boston, MA

BA, Biology major — May 2005
Hamilton College, Clinton, NY

Research Experience:

Doctoral Thesis — Feb 2017-present
Genetics of nontraditional glycemic biomarkers for type 2 diabetes, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

- Developed a dissertation proposal and delivered a proposal seminar to the department
- Currently performing genetic analyses on nontraditional glycemic biomarkers for type 2 diabetes (fructosamine, glycated albumin, 1,5-AG) including SNP heritability estimation and exome sequencing analysis

Glaucoma Genetics Research Coordinator — July 2009-July 2014
Massachusetts Eye and Ear Infirmary/The Channing Division of Network Medicine of Brigham and Women's Hospital and Harvard Medical School, Boston, MA

- Worked primarily on genetic epidemiology research of genome-wide association studies (GWAS) of primary open angle glaucoma (POAG)
- Responsible for logistic and linear regression of genetic and environmental risk factors for POAG. Also performed gene-environment interaction analysis as part of the Gene-Environment Association Studies (GENEVA) consortium

- Worked with collaborators at Vanderbilt University and the Harvard School of Public Health to develop novel pathway analysis methodologies
- Assisted in GWAS imputation and performed analyses on imputed data
- Presented posters at conferences, first author on a paper analyzing genetic risk for POAG stratified by gender and type of vision loss, as well as coauthor on multiple publications

Senior Research Assistant Aug 2006-July 2009
Harvard Partners Center for Genetics and Genomics, Cambridge, MA

- Extensive experience in troubleshooting and optimizing techniques for PCR and ABI 3730xl sequencing
- Helped develop new clinical genetic tests for various diseases including hypertorphic and dialated cardiomyopathy
- Experience with Illumina Genome Analyzer, operating the machine, developing new applications and specializing in library construction including SAGE, plasmid and PCR product libraries
- Gained leadership experience as a senior research assistant by assisting in interviewing, hiring and training
- Took initiative to improve and streamline lab operations and protocols

MPH Applied Learning Experience Sept 2007-May 2008
Extremely discordant sibpair study to determine epidemiological risk factors for progression to advanced age-related macular degeneration, Tufts University School of Medicine, Massachusetts Eye and Ear Infirmary (MEEI), Boston, MA

- Worked closely with an investigator at MEEI to analyze cohort data using McNemar's test and conditional logistic regression
- Analyzed and drew conclusions of a specific subcohort of the study population, culminating in a presentation and written report

Undergraduate Senior Thesis Sept 2004-May 2005
Gene expression levels of four amino acid permeases, GAP1, AGP1, AGP2 and AGP3 in CAP and cap- yeast cells, Hamilton College, Clinton, NY

- Conducted literature search on the thesis subject, then designed and carried out a year-long research project in a molecular genetics laboratory
- Utilized plasmid preparations, plasmid transformations and protein assays to elucidate the role of the Cap protein in the Ras/cAMP pathway in S. cerevisiae yeast

Select Publications:

Loomis SJ, Maruthur, NM, Li M, Baldridge A, North K, Mei H, Morrison A, Carson A, Pankow J, Boerwinkle E, Scharpf R, Rasmussen-Torvik L, Coresh J, Duggal P, Kottgen A, Selvin E, Genome-wide association study of serum fructosamine and glycated albumin in adults without diagnosed diabetes: results from the Atherosclerosis Risk in Communities Study. Received reviewer comments for publication in Diabetes, currently working on responses to resubmit.

Loomis SJ, Klein AP, Lee KE, Chen F, Bomotti S, Truitt B, Iyengar SK, Klein R, Klein BEK, Duggal P. Variation in RNF149 is associated with nuclear lens opacity. *Ophthalmic Epidemiology*. 2017 Nov 28:1-5.

Lindström S, **Loomis S**, Turman C, Huang H, Huang J, Aschard H, Chan AT, Choi H, Cornelis M, Curhan G, De Vivo I, Eliassen AH, Fuchs C, Gaziano M, Hankinson SE, Hu F, Jensen M, Kang JH, Kabrhel C, Liang L, Pasquale LR, Rimm E, Stampfer MJ, Tamimi RM, Tworoger SS, Wiggs JL, Hunter DJ, Kraft P. A comprehensive survey of genetic variation in 20,691 subjects from four large cohorts. *PLoS One*. 2017 Mar 16;12(3).

Loomis SJ, Chen Y, Sacks DB, Christenson ES, Christenson RH, Rebholz CM, Selvin E. Cross-sectional analysis of AGE-CML, sRAGE and esRAGE with diabetes and cardiometabolic risk factors in a community based cohort. *Clinical Chemistry*. 2017 May;63(5):980-989.

Bailey JN, **Loomis SJ**, Kang JH, Allingham RR, Gharahkhani P, Khor CC, Burdon KP, Aschard H, Chasman DI, Igo RP Jr, Hysi PG, Glastonbury CA, Ashley-Koch A, Brilliant M, Brown AA, Budenz DL, Buil A, Cheng CY, Choi H, Christen WG, Curhan G, De Vivo I, Fingert JH, Foster PJ, Fuchs C, Gaasterland D, Gaasterland T, Hewitt AW, Hu F, Hunter DJ, Khawaja AP, Lee RK, Li Z, Lichter PR, Mackey DA, McGuffin P, Mitchell P, Moroi SE, Perera SA, Pepper KW, Qi Q, Realini T, Richards JE, Ridker PM, Rimm E, Ritch R, Ritchie M, Schuman JS, Scott WK, Singh K, Sit AJ, Song YE, Tamimi RM, Topouzis F, Viswanathan AC, Verma SS, Vollrath D, Wang JJ, Weisschuh N, Wissinger B, Wollstein G, Wong TY, Yaspan BL, Zack DJ, Zhang K, Study EN; ANZRAG Consortium, Weinreb RN, Pericak-Vance MA, Small K, Hammond CJ, Aung T, Liu Y, Vithana EN, MacGregor S, Craig JE, Kraft P, Howell G, Hauser MA, Pasquale LR, Haines JL, Wiggs JL. Genome-wide association analysis identifies TXNRD2, ATXN2 and FOXC1 as susceptibility loci for primary open-angle glaucoma. *Nat Genet*. 2016 Feb;48(2):189-94.

Kang JH, **Loomis SJ**, Rosner BA, Wiggs JL, Pasquale LR. Comparison of Risk Factor Profiles for Primary Open-Angle Glaucoma Subtypes Defined by Pattern of Visual Field Loss: A Prospective Study. *Invest Ophthalmol Vis Sci*. 2015 Apr;56(4):2439-48.

Kang JH, **Loomis SJ**, Wiggs JL, Willett WC, Pasquale LR. A prospective study of folate, vitamin b6, and vitamin B12 intake in relation to exfoliation glaucoma or suspected exfoliation glaucoma. *JAMA Ophthalmol*. 2014 May 1;132(5):549-59.

Kang JH, **Loomis SJ**, Yaspan BL, Bailey JC, Weinreb RN, Lee RK, Lichter PR, Budenz DL, Liu Y, Realini T, Gaasterland D, Gaasterland T, Friedman DS, McCarty CA, Moroi SE, Olson L, Schuman JS, Singh K, Vollrath D, Wollstein G, Zack DJ, Brilliant M, Sit AJ, Christen WG, Fingert J, Forman JP, Buys ES, Kraft P, Zhang K, Allingham RR, Pericak-Vance MA, Richards JE, Hauser MA, Haines JL, Wiggs JL, Pasquale LR. Vascular tone pathway polymorphisms in relation to primary

open-angle glaucoma. *Eye (Lond).* 2014 Mar 7.

Loomis SJ, Kang JH, Weinreb RN, Yaspan BL, Cooke Bailey JN, Gaasterland D, Gaasterland T, Lee RK, Lichter PR, Budenz DL, Liu Y, Realini T, Friedman DS, McCarty CA, Moroi SE, Olson L, Schuman JS, Singh K, Vollrath D, Wollstein G, Zack DJ, Brilliant M, Sit AJ, Christen WG, Fingert J, Kraft P, Zhang K, Allingham RR, Pericak-Vance MA, Richards JE, Hauser MA, Haines JL, Pasquale LR, Wiggs JL. Association of CAV1/CAV2 genomic variants with primary open-angle glaucoma overall and by gender and pattern of visual field loss. *Ophthalmology*. 2014 Feb;121(2):508-16.

Pasquale LR, **Loomis SJ**, Weinreb RN, Kang JH, Yaspan BL, Bailey JC, Gaasterland D, Gaasterland T, Lee RK, Scott WK, Lichter PR, Budenz DL, Liu Y, Realini T, Friedman DS, McCarty CA, Moroi SE, Olson L, Schuman JS, Singh K, Vollrath D, Wollstein G, Zack DJ, Brilliant M, Sit AJ, Christen WG, Fingert J, Kraft P, Zhang K, Allingham RR, Pericak-Vance MA, Richards JE, Hauser MA, Haines JL, Wiggs JL. Estrogen pathway polymorphisms in relation to primary open angle glaucoma: an analysis accounting for gender from the United States. *Mol Vis*. 2013 Jul 12;19:1471-81.

Pasquale LR, **Loomis SJ**, Aschard H, Kang JH, Cornelis MC, Qi L, Kraft P, Hu FB. Exploring genome-wide - dietary heme iron intake interactions and the risk of type 2 diabetes. *Front Genet*. 2013;4:7.

Pasquale LR, **Loomis SJ**, Kang JH, Yaspan BL, Abdrabou W, Budenz DL, Chen T, DelBono E, Friedman DS, Gaasterland D, Gaasterland T, Grosskreutz CL, Lee RK, Lichter PR, Liu Y, McCarty CA, Moroi SE, Olson LM, Realini T, Rhee DJ, Schuman JS, Singh K, Vollrath D, Wollstein G, Zack DJ, Allingham RR, Margaret A. Pericak-Vance MA, Robert N. Weinreb (16), Kang Zhang (16), Michael A. Hauser (10,17), Julia E. Richards (9, 18), Jonathan L. Haines (3), Janey L. Wiggs (1). CDKN2B-AS1 Genotype – Glaucoma Feature Correlations in Primary Open-Angle Glaucoma Patients from the United States. *Am J Ophthalmol*. 2013 Feb;155(2):342-353.e5.

Kang JH, **Loomis S**, Wiggs JL, Stein JD, Pasquale LR. Demographic and geographic features of exfoliation glaucoma in 2 United States-based prospective cohorts. *Ophthalmology.* 2012 Jan;119(1):27-35.

Loomis SJ, Olson LM, Pasquale LR, Wiggs J, Mirel D, Crenshaw A, Parkin M, Rahhal B, Tetreault S, Kraft P, Tworoger SS, Haines JL, Kang JH. Feasibility of High-Throughput Genome-Wide Genotyping using DNA from Stored Buccal Cell Samples. *Biomark Insights*. 2010 May 20;5:49-55.